

1 **Multi-ancestry GWAS of Long COVID identifies**
2 **immune-related loci and etiological links to chronic fatigue**
3 **syndrome, fibromyalgia and depression**

4 **Authors:** Ninad S. Chaudhary, Catherine H. Weldon, Priyanka Nandakumar, 23andMe
5 Research Team, Michael V. Holmes* and Stella Aslibekyan*

6 *co-senior authors

7 **Institution:** 23andMe, Inc

8

9

10

11

12

13 **Corresponding Author:**

14 Stella Aslibekyan PhD

15 23andMe. Inc

16

1 ABSTRACT

2 The etiology of Long COVID is poorly understood despite its estimated global burden of
3 65 million cases. There exists a paucity of genetic studies that can shed light on
4 potential mechanisms leading to Long COVID. Using consented and genotyped data
5 from 23andMe adult research participants, we conducted the largest multi-ancestry
6 meta-analysis of genome-wide association studies of Long COVID across European
7 (42,899 cases, 94,721 controls), Latinx (8,631 cases, 20,351 controls), and
8 African-American (2,234 cases, 5,596 controls) genetic ancestry groups. GWAS of Long
9 COVID identified three genome-wide significant loci (*HLA-DQA1–HLA-DQB*, *ABO*,
10 *BPTF–KPAN2–C17orf58*). Functional analysis of these genes points to underlying
11 immune and thrombo-inflammatory mechanisms. We present evidence of shared
12 genetic architecture (genetic correlation p -value < 0.001) of Long COVID with thirteen
13 phenotypes of similar symptomatology or pathophysiology. We identified potential
14 causal roles from liability to chronic fatigue (Mendelian randomization OR=1.59, 95%
15 CI[1.51,1.66]), fibromyalgia (OR=1.54, 95% CI[1.49,1.60]), and depression (OR=1.53,
16 95% CI[1.46,1.61]) with Long COVID, which replicated in the COVID-19 Host Genetics
17 Initiative data, and which are unlikely to originate from collider bias. These findings can
18 help identify populations vulnerable to Long COVID and inform future therapeutic
19 approaches.

20

21

1 INTRODUCTION

2 Long COVID (post-acute sequelae of SARS-CoV2-infection or post-acute COVID-19
3 syndrome) is a debilitating condition characterized by symptoms persisting for months
4 following the acute infection. Since the global spread of SARS-CoV2, more than 65
5 million people have reported experiencing symptoms that can be classified as Long
6 COVID (Davis et al. 2023). It is estimated that 7.3% of SARS-CoV2 positive individuals
7 and 52% of those hospitalized due to COVID-19 in the US (Robertson et al. 2023;
8 Taquet et al. 2021; O'Mahoney et al. 2023) subsequently develop Long COVID.
9 Individuals experiencing Long COVID predominantly report fatigue, cognitive
10 impairment and breathlessness impacting daily activities (O'Mahoney et al. 2023; Ford
11 et al. 2024; Lopez-Leon et al. 2021).

12

13 Vaccination efforts have been shown to reduce the risk of developing Long COVID
14 symptoms in SARS-CoV2 positive individuals (Català et al. 2024; Robertson et al.
15 2023). However, among individuals with Long COVID, it is not clear whether vaccination
16 attenuates symptoms, making it imperative to continue identifying treatment options for
17 Long COVID (Notarte et al. 2022; Tran et al. 2023). NIH-funded clinical trials are
18 underway for exploring the efficacy of Paxlovid for the treatment of Long COVID. This
19 represents just one potential therapeutic strategy: lack of knowledge of the underlying
20 etiological mechanisms represents an ongoing challenge to developing evidence-based
21 treatment options.

22

1 Genome-wide association studies (GWAS) of acute COVID-19 susceptibility and
2 severity have contributed in establishing the role of immune mechanisms in response to
3 acute SARS-CoV2 infection (Shelton et al. 2021; COVID-19 Host Genetics Initiative
4 2023). In contrast, the genetic evidence identifying biological mechanisms in Long
5 COVID is limited (Ledford 2024). Mechanisms that can lead to Long COVID may
6 include persistence of viral burden, autoimmune response, and inflammation. A recent
7 GWAS of Long COVID by the COVID-19 Host Genetics Initiative (HGI) identified a
8 single risk locus (*FOXP4* gene) for Long COVID at chromosome 6 (Lammi et al. 2023).
9 This locus has previously been linked to acute COVID-19 susceptibility and severity
10 (COVID-19 Host Genetics Initiative 2023), recognised risk factors for Long COVID
11 (Whitaker et al. 2022). However, the mechanisms underlying the *FOXP4* genetic signal,
12 postulated to include lung function, explain only a part of the multi-systemic response
13 that is postulated to occur in Long COVID. Overall, the mechanistic understanding of
14 Long COVID continues to be elusive. Larger-scale GWAS of Long COVID restricted to
15 previously COVID-19 infected individuals are well suited to identifying genetic loci
16 conferring risk of Long COVID, while controlling for susceptibility, and may help inform
17 treatment strategies.

18

19 Epidemiological studies have also identified similarities between phenotypic clusters of
20 Long COVID and other post-viral syndromes and chronic conditions. (Sherif et al. 2023;
21 Davis et al. 2023; Thompson et al. 2023). This posits the question of whether a
22 pre-existing chronic condition can increase the risk of developing Long COVID. By
23 leveraging GWAS data on these conditions, we can further elucidate disease etiology to

1 identify putative causal risk factors that may shed light on mechanistic pathways. The
2 23andMe research cohort provides an opportunity to investigate the genetic architecture
3 of Long COVID and to conduct post-GWAS translational studies such as Mendelian
4 randomization to shed light on potential etiology. Accordingly, using ancestrally-diverse
5 data from the 23andMe database, we conducted GWAS in three ancestries and
6 performed meta-analysis to identify genetic variants associated with Long COVID in a
7 cohort of COVID-19 infected individuals. We further investigated genetic correlations
8 between Long COVID and symptomatologically similar phenotypes and explored their
9 potential causal relevance, including the role of acute COVID-19 severity, through
10 Mendelian randomization.

11

12 **METHODS**

13 ***Study population and data collection***

14 Details of the 23andMe Research Cohort have been previously described (Shelton et al.
15 2021). Self-reported data were obtained from online survey questionnaires on
16 demographics, socio-behavioral characteristics, and health conditions. Data on Long
17 COVID were collected through a survey designed to inform about the symptomatic
18 burden and impact on daily living activities. Participants who had previously reported to
19 have COVID-19 diagnosis or tested positive for SARS-CoV2 were invited via email to
20 participate in the Long COVID survey. Additional respondent characteristics, such as
21 age, sex, pre-existing conditions (high blood pressure, diabetes, chronic kidney disease,
22 coronary heart disease, cardiac arrhythmias, high cholesterol, history of blood clots, and
23 chronic obstructive pulmonary disease), educational attainment, smoking status were

1 queried in the baseline COVID-19 surveys if the data were missing from previously
2 distributed health surveys. Data collected between August 2021 and February 2024
3 were used for the current work (**Figure 1**).

4

5 ***Data availability***

6 The full GWAS summary statistics for the 23andMe discovery data set will be made
7 available through 23andMe to qualified researchers under an agreement with 23andMe
8 that protects the privacy of the 23andMe participants. Datasets will be made available at
9 no cost for academic use. Please visit
10 <https://research.23andme.com/collaborate/#dataset-access/> for more information and to
11 apply to access the data.

12

13 ***Ethics statement***

14 Study participants provided informed consent. The recruitment, informed consent, and
15 collection of online data via surveys were conducted according to our human
16 participants research protocol, which was reviewed and approved by Ethical and
17 Independent Review Services, a private institutional review board
18 (<http://www.eandirreview.com>). As of 2022, E&I Review Services is part of Salus IRB
19 (<https://www.versiticlinicaltrials.org/salusirb>).

20 ***Phenotype definition***

21 In contrast to prior Long COVID GWAS using population controls (Lammi et al. 2023),
22 only participants who were diagnosed with COVID-19 or had tested positive for
23 COVID-19 were included in the main analysis. We first excluded those who had

1 reported to have had acute COVID-19 illness within the last three months for
2 consistency with the widely accepted definition whereby to be eligible to be a Long
3 COVID case, it occurred three months after the acute infection (Soriano et al. 2022).
4 Long COVID was defined using self-reported data on the question: “*Have you*
5 *experienced post-acute COVID-19 syndrome (also called 'Long COVID' or 'Long*
6 *Haulers')?*”. Those who responded “yes” to the question were defined as cases of Long
7 COVID. Participants who responded “No” to the question were defined as controls.
8 The data on impact on daily life activities were obtained among individuals who reported
9 having a COVID-19 diagnosis and symptoms lasting for one month or longer since
10 illness from the question “On average, how much has your persistent COVID-19
11 symptoms impacted your normal activities? If you have fully recovered, please respond
12 based on your experience before you fully recovered”. Participants' responses were
13 recorded on a 5 point Likert scale of “not at all, slightly, moderately, quite a bit,
14 extremely”. Among controls, 16,390 individuals had responded to having some impact
15 on daily activities post acute illness and were excluded from this group. In a sensitivity
16 analysis, we modified the definition for Long COVID cases incorporating the impact of
17 Long COVID symptoms on daily life (referred to as “Long COVID Impact” phenotype)
18 (Lammi et al. 2023; Soriano et al. 2022). Long COVID Impact phenotype was defined as
19 those who self-reported Long COVID, and had moderate to severe impact of Long
20 COVID symptoms on daily living activities. The impact was considered moderate to
21 severe if participants chose the following answer options: moderately, quite a bit, or
22 extremely (N=32,346).

23

1 **Genotyping**

2 DNA extraction and genotyping were performed on saliva samples by Clinical
3 Laboratory Improvement Amendments and Collage of American Pathologists-accredited
4 clinical laboratories of Laboratory Corporation of America. The details of genotyping
5 platforms and imputation quality controls have been previously described (Chaudhary et
6 al. 2024). Approximately 110 million variants were available for genome-wide analyses
7 after the quality control procedures. Genotyped data were further analyzed to determine
8 the ancestry of the participants using a genetic ancestry classification algorithm.
9 Subsequent analyses were performed separately for each ancestry. To limit our
10 analyses to unrelated individuals, we selected participants such that no two individuals
11 shared more than 700cM of DNA identical by descent. In a scenario where a case and a
12 control have at least 700cM of DNA identical by descent, we removed the control from
13 the analytical sample. After excluding unrelated individuals while prioritizing the
14 retention of cases, the analytical cohort for GWAS consisted of 53,764 cases and
15 120,688 controls for Long COVID, and 32,087 cases and 121,298 controls for Long
16 COVID Impact.

17

18 **Imputation of classical HLA Alleles**

19 Imputation of *HLA* alleles was performed using the HIBAG13 R package (Zheng et al.
20 2014). HIBAG is an attribute bagging based statistical method that uses a pre-fit
21 classifier to impute the allelic dosages of *HLA* alleles. This classifier was trained from a
22 large internal database of individuals with known *HLA* alleles and SNP variation within
23 the *HLA* region. We used the default setting of HIBAG and used a multi-population

1 reference panel from 100,000 samples of 23andMe internal data and other external
2 data. Approximately 98% of the tagging SNPs used in HIBAG were genotyped and
3 passed quality control on 23andMe's platform. We imputed allelic dosage for *HLA-A*, *B*,
4 *C*, *DPA1*, *DPB1*, *DQA1*, *DQB1* and *DRB1* loci at the p-group resolution where *HLA*
5 alleles that encode the same protein sequence for the peptide binding domains (exon 2
6 and 3 for class I and exon 2 only for class 2) are grouped together and designated with
7 4 digits (Tian et al. 2017). Similar to SNP imputation, we measured imputation quality
8 using r^2 which ranged from 0.72 to 1.00. Overall, there were 855 imputed classical HLA
9 alleles at four-digit resolution available for analysis.

10

11 **Ancestry determination**

12 Ancestries in the 23andMe database are determined using a classifier algorithm based
13 on analysis of local ancestry (Durand et al. 2014). Phased genotyped data were first
14 partitioned into windows of about 300 SNPs and a support vector machine (SVM)
15 approach was applied within each window to classify individual haplotypes into one of
16 45 worldwide reference populations. The SVM classifications are then fed into a hidden
17 Markov model (HMM) that accounts for switch errors and incorrect assignments, and
18 gives probabilities for each reference population in each window. Finally, we used
19 simulated admixed individuals to recalibrate the HMM probabilities so that the reported
20 assignments are consistent with the simulated admixture proportions. We aggregated
21 the probabilities of the 45 reference populations into six main ancestry categories
22 (European, African-American, Latinx, East Asian, South Asian, Middle Eastern) using a
23 predetermined threshold (Durand et al. 2014). African Americans and Latinx were

1 admixed with broadly varying contributions from Europe, Africa and the Americas. No
2 single threshold of genome-wide ancestry could effectively discriminate between African
3 Americans and Latinx. However, the distributions of the length of segments of
4 European, African and American ancestry are very different between African Americans
5 and Latinx, because of distinct admixture timing between the three ancestral
6 populations in the two ethnic groups. Therefore, we trained a logistic classifier that took
7 the participant's length histogram of segments of African, European and American
8 ancestry, and predicted whether the customer is likely African American or Latinx.

9

10 **STATISTICAL ANALYSIS**

11 ***Multi-ancestral GWAS***

12 Within each ancestral group, we performed logistic regression for Long COVID with
13 additive allelic effects for each variant. The model was adjusted for age, age-squared,
14 sex, and the top five within-ancestry principal components to account for residual
15 population structure and genotyping platform. The genome-wide threshold for significant
16 association for each variant was $< 5.0 \times 10^{-8}$. The summary statistics were corrected for
17 inflation using genomic control when the inflation factor was greater than one and
18 presented using Manhattan plots. We then constructed regions of significant association
19 by identifying all SNPs with $P < 10^{-5}$ within the vicinity of a genome-wide significant
20 association. We defined a distance between neighboring regions of association of at
21 least 250kb to identify a significant locus. The SNP with the smallest P-value within
22 each region was identified as the index variant. We also calculated a credible set of
23 variants within each region based on Bayesian approach to refine the association

1 signals (Wellcome Trust Case Control Consortium et al. 2012). The regional LD plots for
2 each index variant were constructed. The analyses were conducted separately in each
3 genetic ancestry cohort (Europeans, African-Americans, and Latinx). Participants
4 classified as other ancestral groups (South Asian, East Asian, and Middle Eastern) were
5 excluded due to insufficient GWAS sample size. We also conducted logistic regression
6 between *HLA* loci and Long COVID adjusting for above mentioned covariates. Previous
7 studies have shown that *HLA* imputed alleles are associated with viral infections as
8 individuals carrying the same *HLA* allelic dosage recognize the same viral epitope (Tian
9 et al. 2017; Augusto et al. 2023; Debebe et al. 2020). We accordingly determined the
10 associations between *HLA* alleles imputed till four-digit resolution and Long COVID by
11 ancestry in the region of statistically significant *HLA* loci.

12

13 **Meta-Analysis**

14 We then performed a multi-ancestry meta-analysis of above GWAS using the inverse
15 variance fixed effect model. Only variants with minor allele frequency of more than one
16 percent in all ancestries were considered for meta-analysis. The associations were
17 further adjusted for inflation and tested for heterogeneity between genetic ancestry
18 groups using the Cochran's Q test. The effect estimates of index variants on
19 meta-analyses are presented as odds ratios (OR) on an additive allelic scale.

20

21 **Functional annotation of GWAS index variants**

22 To perform variant-to-gene mapping, hypotheses of functionally relevant genes are
23 generated by annotating the index variants. Functional variants for mapping include

1 coding variants (annotated by the Ensembl Variant Effect Predictor (VEP) v109 44),
2 eQTLs, and pQTLs. The eQTL annotation resources consist of a comprehensive
3 collection of standardized variants impacting gene expression in various tissues
4 obtained from publicly available datasets (GTEx Consortium 2020; Qiu et al. 2018;
5 Kettunen et al. 2012; Franzén et al. 2016; Raj et al. 2014; Kerimov et al. 2021;
6 Lappalainen et al. 2013; Koolpe et al. 1985; Craig et al. 2021) and datasets processed
7 by the 23andMe eQTL pipeline. The pQTL annotation resources similarly include a
8 collection of curated protein QTLs from relevant public datasets (Zhang et al. 2022;
9 Ferkingstad et al. 2021).

10 eQTL calling was performed with one of two versions of 23andMe pipelines, depending
11 on the dataset in question (**Supplementary File**). The first pipeline used FastQTL
12 (Ongen et al. 2016) in permutation mode, restricting all tests to variants within a
13 window defined to be 1Mbp up- or downstream of a given gene's transcription start site
14 (TSS) (**Supplementary File Table 1**). Variants tested were single nucleotide
15 polymorphisms with an in-sample MAF $\geq 1\%$, to avoid errors in detection or mapping of
16 larger genetic variants in cross-ancestry comparisons, and models were adjusted for
17 age (if available), sex, probabilistic estimation of expression residuals (PEER) factors
18 (Stegle et al. 2012), genetic PCs, and per-dataset covariates. For each gene, the index
19 variant was identified by the minimal permutation p-value. eQTLs were called then on
20 lead variants if they passed a 5% false discovery rate (FDR) filter using Storey's q-value
21 (Storey 2002) methodology. Conditional eQTLs were identified via FastQTL's
22 permutation mode, by using each eQTL as an additional covariate in the model for a
23 given gene. The lead conditional eQTL for all genes were again FDR controlled at 5%,

1 and a maximum of 10 conditional steps were run. Finally, for a set of conditional eQTLs
2 for a given gene, a joint model was fit, and the final eQTL callset consisted of those
3 eQTLs whose joint model test passed a 5% FDR filter. eQTLs were only called for
4 genes classified as one of 'protein_coding', 'miRNA', 'IG_C_gene', 'IG_D_gene',
5 'IG_J_gene', 'IG_V_gene', 'TR_C_gene', 'TR_D_gene', 'TR_J_gene', 'TR_V_gene' as
6 defined in GENCODE (Harrow et al. 2012).

7 The second version of the 23andMe pipeline uses strand-aware RNA-seq quantification,
8 and the eQTLs were called using SusieR package (Wang et al. 2020) instead of
9 FastQTL, with expression PCs (selected with the elbow method) replacing PEER
10 factors in the modeling, and using the GENCODE v43 gene model (**Supplementary**
11 **File Table 2**). The pipeline natively generated credible sets with a set probability to
12 contain a SNP tagging the causal variant.

13

14 ***Genetic correlation of Long COVID with phenotypically similar traits***

15 Though the exact biological mechanisms of Long COVID remain to be determined, it is
16 proposed that Long COVID is a consequence of the multi-system effect of acute-phase
17 COVID-19 infection (Thompson et al. 2023). The phenotypic clusters and
18 pathophysiology of Long COVID overlap with post-viral syndromes, dystonia, chronic
19 fatigue/MFS, endothelial dysfunction, and conditions associated with immune
20 dysregulation and coagulopathies. To understand whether these phenotypic and
21 pathophysiological similarities have genetic underpinnings, we obtained the genetic
22 correlations between Long COVID and symptomatologically similar phenotypes in
23 participants of European ancestry using LD score regression (Bulik-Sullivan et al. 2015).

1 We selected phenotypes if the phenotypes were available in the 23andMe database
2 and 1) the symptomatology of the phenotype overlapped with the symptoms of Long
3 COVID (chronic pain, chronic obstructive pulmonary disease, polymyalgia rheumatica,
4 chronic fatigue/myalgic encephalitis, fibromyalgia, Alzheimer's disease, polymyositis,
5 hypersensitivity pneumonitis, lupus, complex regional pain syndrome, asthma); and/or
6 2) the phenotype had known etiological precedent of viral illness (EBV-related illness
7 such as chronic fatigue syndrome/myalgic encephalitis, dysautonomia, multiple
8 sclerosis, Guillain-Barre syndrome, chronic EBV syndrome, immunodeficiency); and/or
9 3) phenotypes that have biological relevance and analytical associations with Long
10 COVID (depression, anxiety, diabetes, arrhythmias such as ventricular fibrillation and
11 ventricular tachycardia, clotting disorders such as arterial thrombosis, body-mass index,
12 lung cancer, myocardial infarction, fibromyalgia, and chronic fatigue) based on literature
13 review (Su et al. 2023; Thompson et al. 2023; Al-Aly, Xie, and Bowe 2021; Davis et al.
14 2023).

15

16 ***Mendelian randomization***

17 Based on the estimates of genetic correlation between several chronic disease
18 phenotypes and Long COVID, we explored whether there was a potential causal role of
19 genetic liability to these conditions and Long COVID by conducting two sample
20 Mendelian randomization (MR) using the *twosampleMR* package (Bowden et al. 2016)
21 in participants of European ancestry. The phenotypes that met all of the following
22 criteria were further considered for MR analysis: 1) more than 5,000 cases in their
23 respective GWAS analysis; 2) genetic correlations met a statistical p-value threshold of

1 0.002 /(0.05/26); 3) heritability of >1% as estimated by LD score regression; 4)
2 predisposing conditions that are associated with Long COVID in prior epidemiological
3 studies (Taylor et al. 2023; Naik et al. 2024; Ursini et al. 2021). Traits that met these
4 four conditions were chronic fatigue, fibromyalgia, and depression. A sub-cohort of
5 non-overlapping samples between each phenotype and Long COVID was derived and
6 GWAS analysis was performed in participants of European ancestry only. A genetic
7 instrument for each phenotype was derived using summary statistics. The index
8 variants from GWAS analysis (defined as above) that had minor allele frequency of
9 ≥ 0.01 were used to derive a genetic instrument for each phenotype. The estimates for
10 MR instruments were subsequently obtained from Long COVID GWAS conducted in
11 participants of European ancestry. We then fitted random effects inverse variance
12 weighted (IVW) models to obtain MR estimates for each non-COVID-19 phenotype. To
13 test the veracity of the findings, we used Steiger filtering and robust MR approaches
14 including weighted median and MR-Egger methods (Bowden et al. 2016; Bowden,
15 Davey Smith, and Burgess 2015; Hemani, Tilling, and Davey Smith 2017).

16

17 ***Sensitivity analyses***

18 We repeated the above GWAS analytical approach for the Long COVID Impact
19 phenotype to identify genetic signals for a more severe form of Long COVID. Additional
20 sensitivity analyses explored the impact of acute SARS-CoV-2 infection severity (as
21 assessed by hospitalization) because it is a known risk factor for Long COVID. To that
22 end, we first further adjusted the association of index variants and Long COVID from the
23 main GWAS for COVID-19 hospitalization at the time of acute SARS-CoV2 infection.

1 Second, using the same genetic instruments as in the main analysis, we evaluated MR
2 effects of chronic fatigue, fibromyalgia, and depression on COVID-19 hospitalization.
3 We further replicated our MR findings for Long COVID and COVID-19 hospitalization
4 using summary statistics for these phenotypes from COVID-19 Host Genetics Initiative.
5 In case of COVID-19 hospitalization, only multi-ancestry summary statistics for
6 COVID-19 hospitalization (Cases N = 16,512, Controls N = 71321) were available via
7 the HGI website (<https://www.covid19hg.org/>). Here, we first obtained the allele
8 information, and effect estimates for variants in each of our instruments from HGI
9 summary statistics. The variants were mapped using chromosome and position on hg38
10 genome build. Then we harmonized the allele information across 23andMe summary
11 statistics of each phenotype and Long COVID HGI data. The variants that did not align
12 for allele information were excluded from the instrument. We sought to replicate MR
13 findings for depression from two separate external MR instruments for depression
14 obtained using summary statistics from UK Biobank European participants (Howard et
15 al. 2018) and from meta-analysis of three largest studies on depression that included
16 UK BioBank, the 23andMe cohort, and the Psychiatric Genomics Consortium (Howard
17 et al. 2019). The effect estimates for Long COVID were obtained from European GWAS
18 conducted in this study. The depression phenotype in this UK BioBank was defined as
19 self-reported depressive symptoms, or primary or secondary diagnosis of depressive
20 mood disorder(ICD codes: F32—Single Episode Depression, F33—Recurrent
21 Depression, F34—Persistent mood disorders, F38—Other mood disorders and
22 F39—Unspecified mood disorders)(Howard et al. 2018, 2). The summary statistics and
23 MR instruments were accessed from publicly available GWAS catalog and

1 "MRCIEU/MRInstruments" R package respectively. We could not identify external
2 instruments for chronic fatigue and fibromyalgia because none of the existing GWAS
3 studies was sufficiently powered to identify genome-wide significant hits (Hajdarevic et
4 al. 2022; Zorina-Lichtenwaller et al. 2023; Backman et al. 2021) .

5

6 **RESULTS**

7 The selection of the participants is represented in **Figure 1**. Among 332,638 research
8 participants who were eligible to participate in the Long COVID survey and reported
9 acute SARS-CoV2 infection, 64,384 participants reported to have experienced Long
10 COVID and 178,537 participants did not. After exclusions as described in **Figure 1**, we
11 had an analytical cohort of 54,390 cases and 124,777 controls. Characteristics for
12 unrelated cases and controls are summarized in **Table 1**, while characteristics by
13 ancestry group are tabulated in **Supplementary Table 1**. Most participants (78.9%) in
14 the analytical cohort were of European ancestry, followed by Latinx (16.6%) and
15 African-American (4.5%) ancestry. Cases were more likely to be females (66.7%),
16 non-tobacco users (37%), with a median age of 44 years [IQR=24] and mean BMI of
17 27.5 (SD=7.6). Compared to controls, cases were more likely to have high blood
18 pressure (OR= 1.52, 95% CI: 1.48, 1.56), depression (OR= 1.98, 95% CI: 1.92, 2.02) ,
19 cardiometabolic conditions (OR= 1.60, 95% CI: 1.56, 1.64), and autoimmune conditions
20 (OR= 1.55, 95% CI: 1.51, 1.59) after adjusting for age, sex, and ancestry. Among Long
21 COVID cases, 4.7% (vs 0.005% controls) were hospitalized during acute SARS-CoV2
22 infection and 8.0% (vs 0.01% controls) had severe respiratory disease. When asked

1 about the impact of Long COVID symptoms on daily life, 57.3% of Long COVID cases
2 reported moderate to severe impact.

3

4 GWAS of Long COVID

5 Using approximately 110 million imputed genetic variants, we identified three loci
6 associated with Long COVID in multi-ancestral analysis based on data from European,
7 African-American, and Latinx ancestry groups (**Figure 2, Table 2**).

8

9 *Genome-wide significant loci*: The top GWAS significant hit, (rs9273363; A/C with C
10 being the effect allele, OR= 1.06, 95% CI: 1.04,1.08, $p = 3.79 \times 10^{-11}$) was located in
11 chromosome 6 in the intergenic region spanning *HLA-DQA1* and *HLA-DQB1*
12 (**Supplemental Figure 1**). The effect sizes for rs9273363 were similar across all three
13 ancestries (p -value for heterogeneity across ancestries= 0.75). Further analysis of
14 specific *HLA* alleles and Long COVID showed that *HLA-DRB1*11:04* (OR=1.18, 95%
15 CI: 1.11, 1.24, $p = 7.0 \times 10^{-09}$), *HLA-C*07:01* (OR=0.94, 95%CI = 0.92, 0.96, $p = 1.14 \times$
16 10^{-08}), *HLA-B*08:01* (OR=0.93, 95%CI = 0.91, 0.95, $p = 1.5 \times 10^{-08}$), and
17 *HLA-DQA1*03:01* (OR=0.95, 95% CI = 0.93, 0.97, $p = 4.1 \times 10^{-08}$) were significantly
18 associated with Long COVID. The effect sizes were similar across three ancestries
19 (**Supplementary Figure 2**).

20

21 The second most significant variant was observed at chromosome 9 in the ABO gene
22 (rs644234, G/T with T being the effect allele, OR= 0.95, 95% CI: 0.94, 0.97, $p = 3.66 \times$
23 10^{-09} , p -value for heterogeneity across ancestries= 0.56) (**Supplementary Figure 3**).

1 Specifically, rs644234 is in linkage disequilibrium (LD) with a frame-shift variant
2 (rs8176719) within the ABO gene (European $r^2= 0.97$, African-American $r^2= 0.44$,
3 Latinx $r^2= 0.90$). rs8176719 is one of the main variants determining the ABO blood
4 group and has been previously linked to COVID-19 susceptibility and severity (Bugert et
5 al. 2012; Severe Covid-19 GWAS Group et al. 2020; Shelton et al. 2021).

6
7 Another association was observed with rs2080090 (A/T with T being the effect allele;
8 OR= 0.95, 95% CI: 0.93, 0.97, $p= 1.36 \times 10^{-08}$; p -value for heterogeneity across
9 ancestries=0.69) located in vicinity of the *BPTF* gene on chromosome 17
10 (**Supplementary Figure 4**). There are at least 3 plausible causal genes in the locus;
11 *BPTF*, *KPAN2*, and *C17orf58*. We observed eQTL signals in lymphoblast cells for a
12 variant in LD (rs12601921, $r^2=0.98$, $p= 4.31 \times 10^{-89}$) with rs2080090 for the *BPTF* gene
13 (**Supplementary Table 2a**). rs2080090 was also in LD with a coding variant, rs7502307
14 ($r^2=0.82$), located at *C17orf58* (**Supplementary Table 2b**).

15
16 In European ancestry participants, the top associated SNP in this locus (rs78794747,
17 A/T with T being the effect allele) was mapped to the intergenic region of *C17orf58* and
18 *KPAN2*. rs78794747 is in high LD with multiple variants associated with gene
19 expression in CD14 monocytes, frontal cortex, and islets of Langerhans cells
20 (**Supplementary Table 2c**). The *C17orf58* gene has been associated with posterior
21 myocardial infarction (Norland et al. 2019). *KPAN2* encodes nuclear transport factor
22 importin $\alpha 1$, which is associated with viral suppression (Miyamoto et al. 2021). Previous
23 studies have shown that an accessory protein encoded by the SARS-CoV2 genome,

1 ORF6, binds to importin α 1, enhancing viral propagation by inhibiting interferon type 1
2 signaling (International HIV Controllers Study et al. 2010; Yuen et al. 2020; Miyamoto et
3 al. 2021). Ancestry-specific GWAS hits are described further in **Supplemental Results**
4 and **Supplementary Table 3a-3c**.

5

6 Associations between genome-wide significant variants and Long COVID persisted
7 upon adjusting for COVID-19 hospitalization (**Supplementary Table 4a**). When we
8 assessed the associations between these variants and COVID-19 hospitalization in our
9 study adjusting for age, age*sex, and principal components of ancestry, none of them
10 was significantly associated (**Supplementary Table 4b**).

11

12 In the summary statistics of the HGI Long COVID GWAS including 6,450 Cases and
13 46,208 controls, effect sizes of our index variants were similar. However, none of the
14 associations observed in our study was statistically significant in HGI data, which
15 consisted of almost 10x fewer cases (**Supplementary Table 5**). When we examined the
16 effect of rs9367106 near *FOXP4* (the significantly associated variant in HGI analyses of
17 Long COVID cases and population controls), the association in our study was not
18 statistically significant ($p=0.57$). Consistently with our observation, this variant was not
19 significantly associated with Long COVID in the HGI data ($p=0.04$) when the analyses
20 were limited to individuals who reported to be positive for SARS-CoV2.

21

22 *GWAS of Long COVID Impact*

1 In addition to *HLA-DQA1* and *HLA-DQB1* mentioned above, we identified one indel
2 variant on chromosome 6 in the intergenic region of *MICA* and *MICB* (OR = 0.91, 95%
3 CI: 0.88,0.94, $p = 1.18 \times 10^{-09}$, p-heterogeneity across ancestries=0.61)
4 (**Supplementary Figure 5, Supplementary Table 6**). The variant, rs9281499, was in
5 LD with multiple variants that showed statistically significant pQTL signals for *MICB*
6 protein in blood plasma (**Supplementary Figure 6, Supplementary Table 7a**).
7 Additionally, rs9281499 is in LD with rs2596510 ($r^2=0.82$) which is associated with
8 *CYP21A2* expression in type1 ($p=2.09 \times 10^{-10}$) and type 2 ($p=2.96 \times 10^{-13}$) T-helper cells,
9 and CD4 T lymphocytes (2.45×10^{-10}) (**Supplementary Table 7b**).

10

11 There were two novel statistically significant loci in the multi-ancestry GWAS of Long
12 COVID Impact phenotype. The top hit, rs58970548, is an indel variant located at the
13 intergenic region of *NMUR2* and *GRIA1* (OR= 0.94, 95% CI: 0.93, 0.96, $p=4.27 \times 10^{-8}$,
14 p-heterogeneity across ancestries=0.06). (**Supplementary Figure 7**). An additional
15 significant locus was observed at rs190759626 (G/T with T being the effect allele)
16 located at the intergenic region of *NUTM2A* and *NUTM2D* (**Supplementary Figure 8**)
17 on chromosome 10. Ancestry specific analyses are reported in **Supplementary**

18 **Results.**

19

20 Genetic Correlation with Similar Phenotypes

21

22 We identified 26 conditions in participants of European ancestry that met the criteria of
23 shared pathophysiology, similar clinical features, and previously published associations
24 with Long COVID **Supplementary Table 8** (Sherif et al. 2023; Davis et al. 2023; Peter
25 et al. 2022; Cohen et al. 2022; Bull-Otterson et al. 2022; Davis et al. 2021; Ursini et al.

1 2021). We further limited the list of conditions to those with at least 5,000 cases
2 contributing to GWAS, p -value < 0.002 ($0.05/26$) for the genetic correlation with Long
3 COVID, and estimated heritability more than 1%. Overall, 13 phenotypes met the
4 criteria and are presented in **Figure 3**. The strongest genetic correlations were
5 observed between Long COVID and chronic pain ($r_g=0.72, p=2.49 \times 10^{-170}$), chronic
6 fatigue ($r_g=0.71, p=1.09 \times 10^{-120}$), and fibromyalgia ($r_g=0.70, p=6.30 \times 10^{-142}$).

7

8 Mendelian Randomization

9 We investigated the potential causal role of three phenotypes (chronic fatigue,
10 fibromyalgia, and depression) with Long COVID through Mendelian randomization
11 (MR). We did not consider chronic pain due to its nonspecific definition in the 23andMe
12 database. The genetic instrument for chronic fatigue included 186 SNPs (mean
13 F-statistic per SNP = 43.17), for fibromyalgia included 349 SNPs (mean F-statistic per
14 SNP=45.04), and for depression included 696 SNPs (mean F-statistic per SNP=53.8).
15 We found strong evidence of a potential causal effect to each of these three conditions
16 on Long COVID (chronic fatigue: OR=1.59 (95%CI: 1.51, 1.66), fibromyalgia: OR=1.54
17 (95%CI: 1.49, 1.60), and depression: OR=1.53 (95%CI: 1.46, 1.61); estimates from
18 IVW-MR). These effects persisted when employing robust MR approaches including
19 weighted median and MR Egger (**Figure 4, Supplementary Figure 9, Supplementary**
20 **Table 9a**). Steiger filtered and outlier filtered estimates (**Supplementary Table 9a**) were
21 similar to the respective IVW estimates for chronic fatigue and fibromyalgia. However,
22 for depression, the Steiger filtered estimates were closer to the null than IVW estimates.
23 The MR-Egger intercept for fibromyalgia was non-zero ($-0.01, p = 0.01$) with the

1 resultant MR-Egger causal estimate being stronger than IVW (**Supplementary Table**
2 **9a**). The Mendelian randomization findings were replicated and were in a direction
3 similar to our findings when the genetic instruments for the exposures derived from the
4 23andMe research cohort were applied to disease estimates derived from the Long
5 COVID HGI consortium (**Supplementary Figure 10, Supplementary Table 10a**). In the
6 Mendelian randomization analyses, the directions of effects for exposures applied to
7 COVID-19 hospitalization data from the COVID HGI consortium were also similar to our
8 findings (**Supplementary Table 10b, Supplementary Figure 11**). When we repeated
9 Mendelian randomization using the Long COVID Impact definition, the results were
10 consistent (**Figure 4, Supplementary Figure 12, Supplementary Table 9b**). MR
11 analysis showed evidence of genetic liability to chronic fatigue syndrome, fibromyalgia,
12 and depression leading to a higher risk of COVID-19 hospitalization (**Figure 4,**
13 **Supplementary Figure 13, Supplementary Table 9c**). Our results were also
14 consistent after using external instruments for depression derived solely from UK
15 Biobank GWAS which included 35 SNPs with a mean F statistic of 40.83, and
16 instruments derived from a meta-analysis of the UK BioBank, the 23andMe cohort, and
17 the Psychiatric Genomics Consortium including 87 SNPs with mean F statistic of 77.67
18 (**Supplementary Figure 14**).

19

20 **DISCUSSION**

21 We report the largest GWAS of Long COVID to date among ancestrally diverse
22 participants with a history of SARS-CoV2 acute infection. Our top finding links the *HLA*
23 region, in particular the *HLA-DRB1*11:04* variant, to developing Long COVID. Additional

1 hits were observed in *ABO*, a previously reported severity locus, and a novel region
2 spanning the *BPTF- C17orf58- KPNA2* genes. Genetic correlation analysis showed
3 strong relationships between Long COVID and multiple chronic conditions, including
4 chronic fatigue, depression, and fibromyalgia. We identified strong evidence of a
5 potential causal effect of genetic liability to these conditions on the risk of Long COVID.
6
7 To date, only one study has reported GWAS findings of Long COVID, conducted by the
8 COVID-19 Host Genetics Initiative (HGI) (Lammi et al. 2023). The HGI study compared
9 6,450 Long COVID cases with 1,093,995 population controls from multiple countries
10 and identified a genome-wide significant association with rs9367106 in the *FOXP4*
11 locus. In contrast, our study has approximately a 7-fold larger number of cases, and is
12 contained within one cohort, increasing statistical power through sample size and
13 removing between-study heterogeneity, respectively. The association detected in the
14 COVID-19 HGI study was not significant in our analysis; cases and controls in our study
15 were both infected with SARS-CoV2, removing the effects of COVID-19 susceptibility by
16 design. The *FOXP4* locus identified by HGI was functionally mapped to COVID-19
17 severity and lung function (Lammi et al. 2023; Wu et al. 2021; Pairo-Castineira et al.
18 2023), and potentially reflects the genetic signal relevant to acute infection susceptibility
19 and/or severity. In contrast, our analysis identified multiple loci (*HLA-DRB1*–
20 *HLA-DQA1*, *BPTF*) that are independent of SARS-CoV-2 hospitalization, another
21 marker of COVID-19 severity, and instead are enriched for immune pathways.
22 Assessing the genome-wide significant hits of our study in the HGI data, the direction
23 and effect sizes for these variants were similar but did not meet statistical significance,

1 likely because our study was powered to pick up more subtle signals (**Supplemental**
2 **Table 5**).

3

4 We observed associations of Long COVID with variants in the *HLA* locus, highlighting
5 the role of immune mechanisms as postulated in earlier studies (Klein et al. 2023; Davis
6 et al. 2023; Augusto et al. 2023). *HLA* class II molecules including our top hit,
7 *HLA-DRB1*, play a critical role in the adaptive immune system by presenting foreign
8 peptide antigens to receptors on T lymphocytes, initiating an immune response on
9 exposure to viral antigens. Specifically, *HLA-DRB1*11* alleles have been previously
10 associated with asymptomatic SARS-CoV2 infections (Augusto et al. 2023; Ouedraogo
11 et al. 2023). While the role of *HLA* variation in Long COVID has not been extensively
12 studied, a prior smaller immunoprofiling study has found monocyte expression of MHC
13 class II (HLA-DR) to be higher among Long COVID cases compared to controls who
14 had experienced an acute SARS-CoV2 expression and recovered (Klein et al. 2023).
15 The discovery of the *HLA* association in a hypothesis-free analysis underscores the
16 importance of antigen presentation in the genetic susceptibility to Long COVID,
17 potentially informing future risk stratification and treatment approaches.

18

19 Our study contributes to the body of evidence in support of the role of *ABO* loci in
20 SARS-CoV2 infection outcomes. Previous studies have identified *ABO* loci as
21 independently associated with COVID-19 susceptibility and severity (Shelton et al.
22 2021; Severe Covid-19 GWAS Group et al. 2020; Pairo-Castineira et al. 2023; Zhao et
23 al. 2021). While differential susceptibility to SARS-CoV2 infection was controlled for in

1 the design of our study, acute infection severity is a well-established risk factor for long
2 COVID (Tsampasian et al. 2023). The association between *ABO* variation and
3 COVID-19 outcomes is potentially driven by the *ABO* antibodies interfering with the
4 attachment of viral spike proteins to ACE2 receptors (Breiman, Ruvèn-Clouet, and Le
5 Pendu 2020; Guillon et al. 2008). Additionally, a proteomic analysis of rs657152 (LD r^2
6 = 0.99 with our GWAS hit rs644234) within the *ABO* locus in a population-based cohort
7 identified proteins related to immune function such as IL-15 and STAT3 signaling
8 (Steffen et al. 2022). IL-15 has been proposed as a therapeutic strategy to improve
9 acute COVID-19 outcomes (Lu et al. 2022; National Library of Medicine (U.S.) 2020;
10 Kandikattu et al. 2020). Blood group subtypes encoded by *ABO* are also associated
11 with differential risk of thrombosis (Vasan et al. 2016), which represents another
12 potentially relevant etiologic mechanism linking *ABO* and Long COVID. In recent
13 studies, thrombo-inflammation has been reported as a clinical feature of Long COVID,
14 likely driven by dysregulated coagulation mechanisms in response to viral
15 antigen-antibody complexes (Cervia-Hasler et al. 2024).

16

17 We provide novel evidence establishing potential causal relationships between Long
18 COVID and genetically correlated conditions that share similar symptomologies, namely
19 chronic fatigue, depression, and fibromyalgia. The findings that genetic liability to
20 chronic fatigue, depression and fibromyalgia also leads to a higher risk of hospitalization
21 for acute SARS-CoV-2 infection provides mechanistic insights. Prior studies have
22 reported depression can impair function of a broad range of immune cells (Drevets et al.
23 2022; Walther et al. 2022). While the relationship of chronic fatigue and fibromyalgia

1 with immune dysfunction is less clear, NK cell activity may be impaired among these
2 individuals (Brenu et al. 2012; Rivas et al. 2018; Verma et al. 2022), with altered NK cell
3 function also identified among patients with acute severe COVID-19 (Malengier-Devlies
4 et al. 2022). Furthermore, there is growing evidence that immune and oxidative stress
5 pathways induce neuroaffective effects resulting in symptoms such as chronic fatigue
6 syndrome and depression (Al-Hakeim et al. 2023). A recent study found that Long
7 COVID and chronic fatigue syndrome share a common picture of impaired immune
8 response secondary to metabolic consequences of elevated oxidative stress in T
9 lymphocytes (Shankar et al. 2024). Thus the immune impact of chronic fatigue,
10 depression, and fibromyalgia potentially explains the well-documented higher risk of
11 Long COVID among individuals who experienced severe acute COVID-19. While the
12 immune pathway likely doesn't fully account for the increased risk of Long COVID, it
13 provides new information that could help identify subgroups of individuals that ought to
14 receive preventative measures such as ongoing vaccination to avoid severe acute
15 infection. Also notable from a translational perspective is the indel variant, rs58970548,
16 near the *GRIA1* gene in the GWAS of the Long COVID impact phenotype. Though we
17 could not identify any functional signals at predetermined thresholds of significance
18 ($r^2 > 0.80$ and $p < 5 \times 10^{-08}$) in our datasets, the variants within 500kb and in linkage
19 disequilibrium ($r^2 > 0.5$) of rs58970548 have been associated with various psychiatric
20 conditions (Parekh et al. 2018; Ismail et al. 2022; Chiesa et al. 2012). Specifically,
21 *GRIA1* has been recently identified in brain proteome data as a potential drug target for
22 depression (Liu et al. 2023).

23

1 Our study has multiple strengths and limitations. First, this is the largest GWAS study of
2 Long COVID to date that identified multiple etiologically relevant loci that help shed light
3 on Long COVID biology. Given the constellation of symptoms of Long COVID and their
4 varying severity at individual level (Davis et al. 2021), setting clinical trials for Long
5 COVID has been challenging in the absence of such data on biological mechanisms. By
6 providing evidence of immune function being a likely driving force for etiological tapestry
7 of Long COVID and associated conditions, there is an opportunity to reassess the
8 feasibility of repurposing existing therapeutic interventions planned in the Long COVID
9 trials. Second, the extensive database of 23andMe provided a unique opportunity to
10 study the shared genetic architecture of multiple phenotypes, linking Long COVID with
11 other disorders characterized by substantial unmet clinical need such as chronic fatigue
12 syndrome, depression, and fibromyalgia. Third, owing to the inclusion of more
13 genome-wide significant SNPs compared to prior studies (Lammi et al. 2023), the
14 genetic instruments explain a greater proportion of the genetic variability in the
15 exposures, potentially leading to more reliable MR estimates. Our MR estimates were
16 successfully replicated in an external dataset, lending further credence to our causal
17 conclusions. However, our findings must be interpreted in the context of several
18 limitations. The use of self-reported data enabled us to capture cases that may have
19 been overlooked due to diagnostic challenges surrounding Long COVID (Knuppel et al.
20 2023). Nevertheless, the self-reported data may also have introduced misclassification
21 bias due to subjectivity in grading of symptoms at the individual level. Additionally,
22 because participation in 23andMe surveys is self-selected, our findings may not be
23 representative of the broader population. Furthermore, collider bias is a recognised

1 threat to the validity of COVID-19 studies (Griffith et al. 2020). However, the 696 genetic
2 variants comprising the MR instrument for depression were not associated with
3 participation into our study ($p= 0.82$) and almost all study participants responding to a
4 survey reported having experienced acute COVID-19, making both study respondents
5 to the Long Covid survey and study participation on the basis of previous acute
6 COVID-19 infection unlikely colliders, collectively alleviating selection bias concerns.

7

8 In summary, the largest multi-ancestry GWAS of Long COVID to date highlights new
9 biological mechanisms independent of COVID-19 susceptibility. Furthermore, we report
10 robust genetic correlations between Long COVID and a number of phenotypes with
11 similar symptoms, including chronic fatigue, fibromyalgia, and depression, as well as a
12 causal effect of liability to these conditions on Long COVID. Together, these findings can
13 help identify at-risk individuals for Long COVID, as well as provide novel insights that
14 support developing therapeutic options for both Long COVID and symptomatically
15 similar conditions.

16

17

18

19

20

21

22

23

1 **ACKNOWLEDGEMENTS**

2 We would like to thank the research participants and employees of 23andMe for making
3 this work possible. The following members of the 23andMe Research Team contributed
4 to this study:

5 Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg,
6 Ninad S. Chaudhary, Zayn Cochinwala, Sayantan Das, Emily DelloRusso, Payam
7 Dibaeinia, Sarah L. Elson, Nicholas Eriksson, Chris Eijsbouts, Teresa Filshtein, Pierre
8 Fontanillas, Davide Foletti, Will Freyman, Zach Fuller, Julie M. Granka, Chris German,
9 Éadaoin Harney, Alejandro Hernandez, Barry Hicks, David A. Hinds, M. Reza
10 Jabalameli, Ethan M. Jewett, Yunxuan Jiang, Sotiris Karagounis, Lucy Kaufmann, Matt
11 Kmiecik, Katelyn Kukar, Alan Kwong, Keng-Han Lin, Yanyu Liang, Bianca A. Llamas,
12 Aly Khan, Steven J. Micheletti, Matthew H. McIntyre, Meghan E. Moreno, Priyanka
13 Nandakumar, Dominique T. Nguyen, Jared O'Connell, Steve Pitts, G. David Poznik,
14 Alexandra Reynoso, Shubham Saini, Morgan Schumacher, Leah Selcer, Anjali J.
15 Shastri, Jingchunzi Shi, Suyash Shringarpure, Keaton Stagaman, Teague Sterling,
16 Qiaojuan Jane Su, Joyce Y. Tung, Susana A. Tat, Vinh Tran, Xin Wang, Wei Wang,
17 Catherine H. Weldon, Amy L. Williams, Peter Wilton.

18

19

20

21

22

23

24

25

26

27

28

1

2 **AUTHOR INFORMATION**

3 ***Affiliations:***

4 23andMe, Inc:

5 Ninad S. Chaudhary, Catherine H. Weldon, Priyanka Nandakumar,

6 23andMe Research Team, Michael V. Holmes, Stella Aslibekyan

7

8 ***Author Contributions:***

9 The 23andMe COVID-19 Team developed the recruitment and participant engagement
10 strategy and acquired and processed the data. N.S.C analyzed the data. N.S.C., P.A,
11 C.H.W, S.A, and M.V.H interpreted the data. N.S.C., S.A. and M.V.H. wrote the
12 manuscript. All authors participated in the preparation of the manuscript by reading and
13 commenting on the drafts before submission.

14

15

16

17

18

19

20

21

22

23

24

1 ETHICS DECLARATIONS

2 *Competing Interests*

3 C.H.W, P.N, S,A, and M.V.H are current employees of 23andMe and hold stock or stock
4 options in 23andMe. N.S.C works as a postdoctoral fellow on the 23andMe Genetic
5 Epidemiology Team.

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

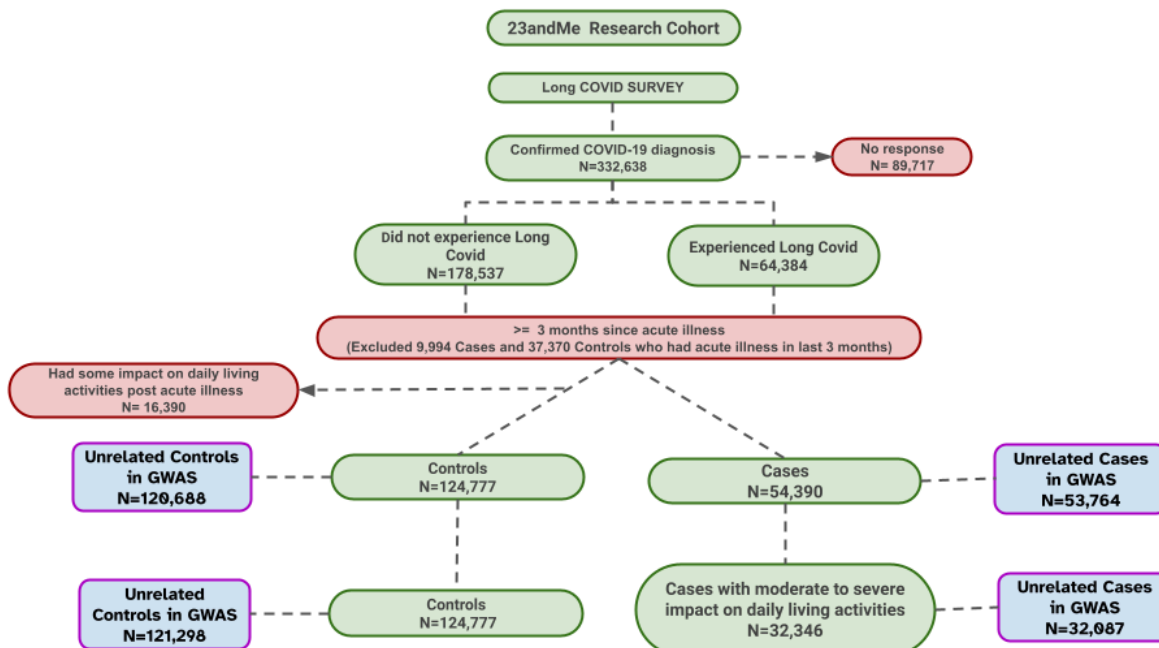
37

1 **FIGURE LEGENDS:**

2

3 **Figure 1: Flowchart of 23andMe participants who participated in Long COVID**
 4 **survey and were included in the analysis**

5 47,364 participants were excluded if the symptoms are reported less than 3 months
 6 from acute illness; 16,390 participants were excluded from the control group if they
 7 reported some impact (selected option other than “not at all”) on daily living activities
 8 post acute illness. Participants of other ancestries (East Asian, South Asian, and Middle
 9 Eastern) present in 23andMe research database were not included in the base cohort
 10 due to insufficient sample size.

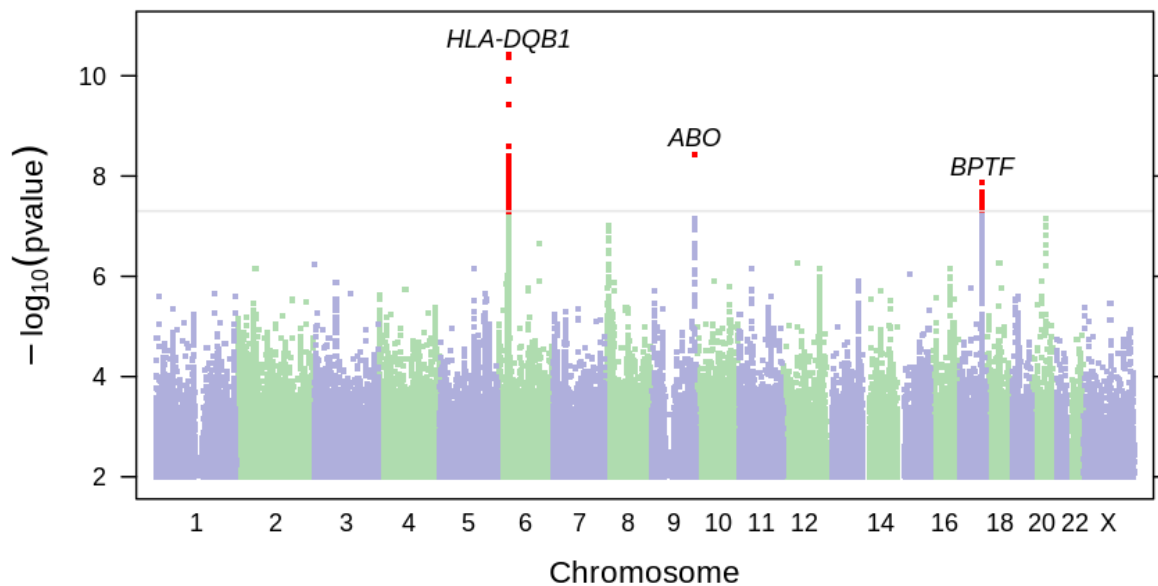


11
12
13
14
15
16
17
18
19
20
21
22
23
24

1 **Figure 2: Manhattan plot of Long COVID among 23andMe participants**

2 Manhattan plot depicts findings from the meta-analysis of three ancestral groups
3 (European, African American, and Latinx). X-axis represents chromosomal position for
4 each SNP. Y-axis represents negative log p-values based on logistic regression model
5 under the additive model. Statistically significant variants are highlighted in red. The
6 regions of associations are annotated with index variants.

7



8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

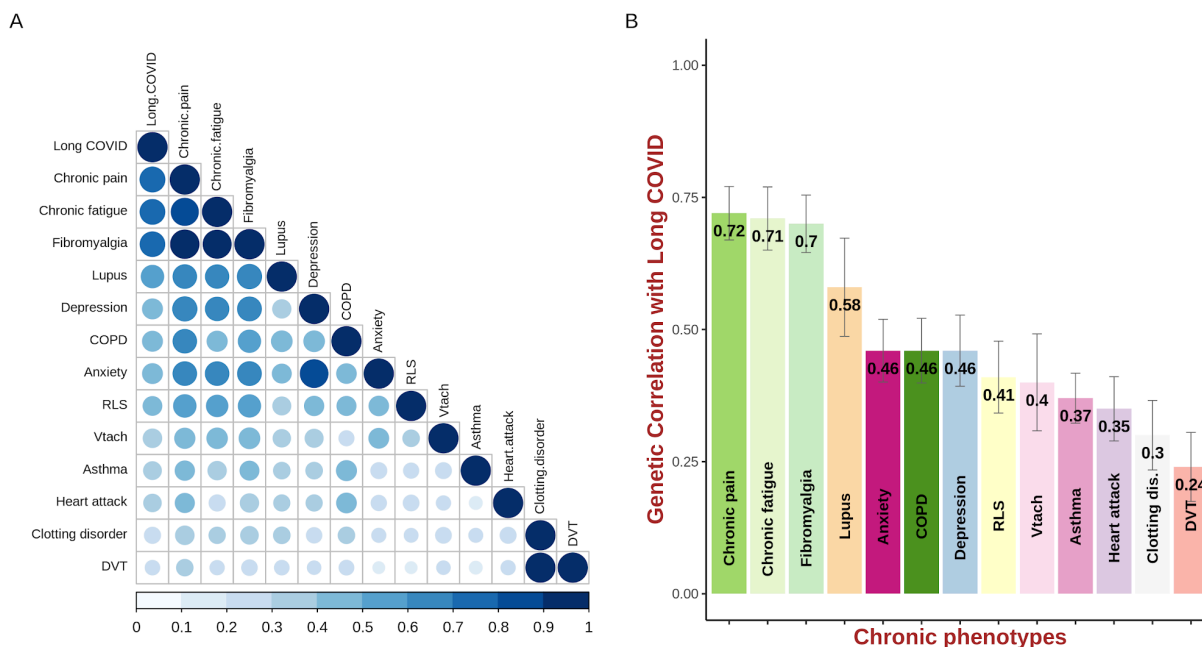
1 Figure 3: Genetic correlation of Long COVID with chronic phenotypes in 23andMe
2 participants of European ancestry

3 Panel A: Heat map showing genetic correlations between chronic phenotypes,
4 including Long COVID, obtained from LD Score regression.

5 Panel B: Bar plot represents genetic correlation values of chronic phenotypes and Long
6 COVID obtained from LD Score regression. X-axis shows a list of chronic phenotypes
7 plotted. Y-axis shows values of genetic correlation from 0 to 1. A complete list of
8 phenotypes are included in Supplementary Table 8.

9 RLS = Restless leg syndrome; Vtach = Ventricular tachycardia; DVT = Deep vein
10 thrombosis; dis. = disorders; COPD = Chronic obstructive pulmonary disease

11



12

13

14

15

16

17

18

19

20

21

22

23

24

1 Figure 4: Forest plot representing genetically predicted effects of chronic
2 phenotypes using Mendelian randomization

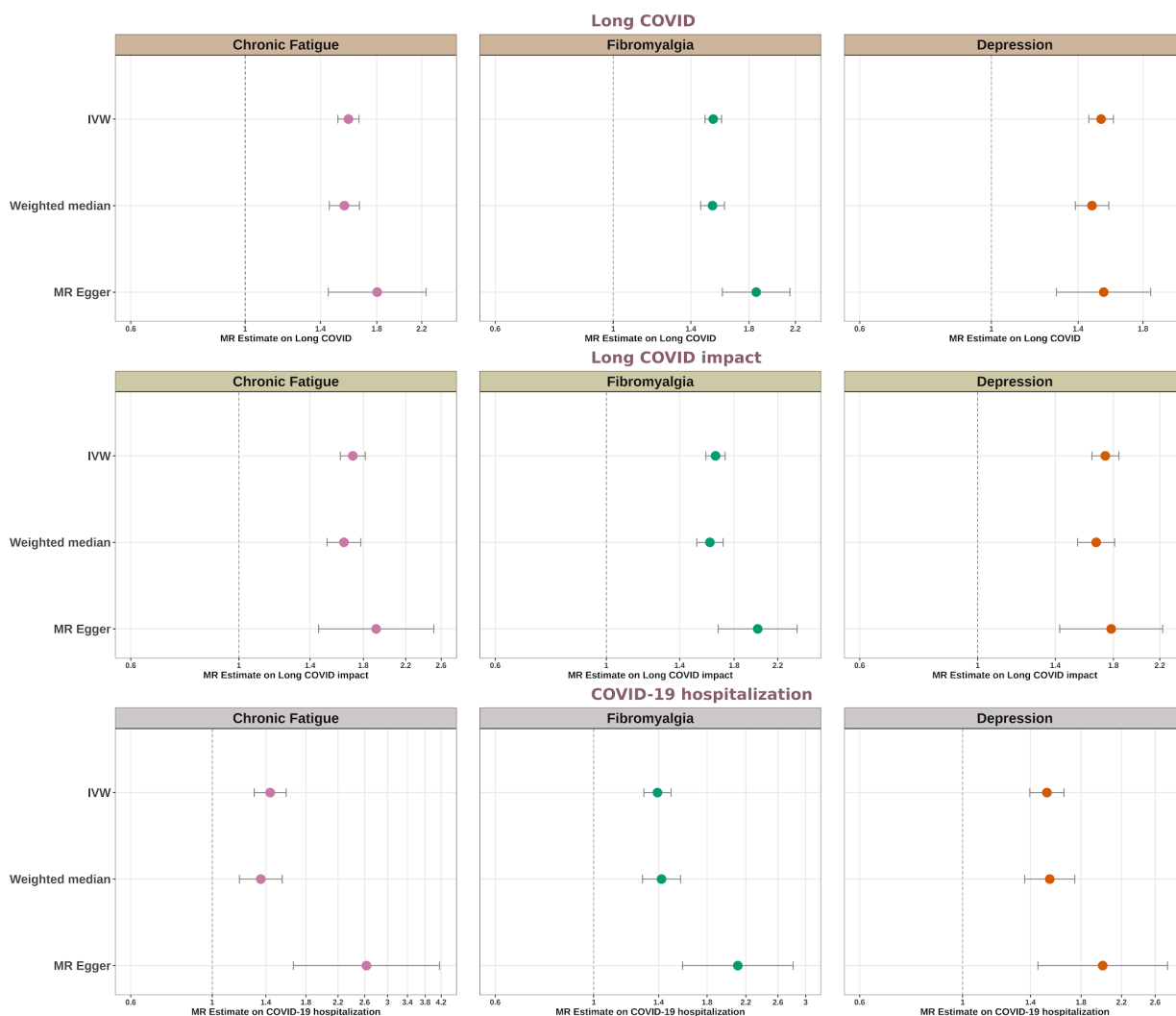
3 Top row panel: Genetically predicted effects of chronic phenotypes on Long COVID

4 Middle row panel: Genetically predicted effects of chronic phenotypes on Long COVID
5 impact

6 Bottom row panel: Genetically predicted effects of chronic phenotypes on COVID-19
7 hospitalization

8 The estimates in the plot depict odds ratio and 95% confidence intervals. The genetic
9 instrument for chronic fatigue is derived using information from 186 SNPs (mean F
10 statistic = 43.17). The genetic instrument for chronic fatigue is derived using information
11 from 349 SNPs (mean F statistic = 45.04). The genetic instrument for depression is
12 derived using information from 696 SNPs (mean F statistic = 53.8). IVW = Inverse
13 variance weighted.

14



15

16

1
2

Table 1: Characteristics of 23andMe Research participants of Long COVID cohort		
	CASE	CONTROL
N	53,764	120,668
Demographics		
Age (median [IQR])	46.0[35.0,57.0]	43.0[33.0,57.0]
Female N (%)	41,324 (76.9)	74,981 (62.1)
Race N (%)		
African American	2,234 (4.2)	5,596 (4.6)
European	42,899 (79.8)	94,721 (78.5)
Latino	8,631 (16.1)	20,351 (16.9)
Education in years (mean (SD))	15.1 (2.4)	15.4 (2.6)
BMI (median [IQR])	28.4 [24.3,34.0]	27.3 [23.8,31.8]
Tobacco user N (%)	21,223 (39.5)	41,921 (35.7)
Two or more alcoholic drinks over past 2 weeks N (%)	10,114 (18.8)	28,733 (23.8)
Health Conditions		
High blood pressure N (%)	14,977 (27.8)	25,507 (21.1)
Depression N (%)	25,226 (46.9)	37,578 (31.1)
COPD N (%)	911 (1.7)	840 (0.01)
High total cholesterol N (%)	4,487 (8.3)	8,054 (6.7)
High triglyceride N (%)	3,207 (6.0)	5,444 (4.5)
Chronic Fatigue N (%)	3,980 (7.4)	1,866 (1.5)
Fibromyalgia N (%)	4,394 (8.2)	2,631 (2.2)
Cardiometabolic condition N (%)	23,566 (43.8)	41,115 (34.1)
Autoimmune N (%)	9,922 (18.4)	14,053 (11.6)
COVID-19+ and hospitalized N (%)	2,534 (4.7)	617 (0.005)
COVID-19+ and severe respiratory disease N (%)	4,282 (8.0)	1,198 (0.01)

3
4
5
6
7
8
9
10

1
2
3

Table 2: Estimates of genome-wide significant variants in meta-analysis of Long COVID GWAS

SNP	Chr	Position	Gene	Alleles	Effect allele	EAF	OR (95% CI)	P value
rs9273363	chr6	32658495	HLA-DQA1--[HLA-DQB1	A/C	C	0.74	1.06(1.04,1.08)	3.79X10 ⁻¹¹
rs644234	chr9	133266804	[ABO]	G/T	T	0.64	0.95(0.94,0.97)	3.66X10 ⁻⁰⁹
rs2080090	chr17	67832255	[BPTF]	A/T	T	0.77	0.95(0.93,0.97)	1.36X10 ⁻⁰⁸

EAF = Effect allele frequency; OR (95% CI)=Odds ratio with 95% confidence interval; Chr = Chromosome

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

1 REFERENCES

2

3 Al-Aly, Ziyad, Yan Xie, and Benjamin Bowe. 2021. “High-Dimensional Characterization
4 of Post-Acute Sequelae of COVID-19.” *Nature* 594 (7862): 259–64.
5 <https://doi.org/10.1038/s41586-021-03553-9>.

6 Al-Hakeim, Hussein Kadhem, Haneen Tahseen Al-Rubaye, Dhurgham Shihab
7 Al-Hadrawi, Abbas F. Almulla, and Michael Maes. 2023. “Long-COVID Post-Viral
8 Chronic Fatigue and Affective Symptoms Are Associated with Oxidative Damage,
9 Lowered Antioxidant Defenses and Inflammation: A Proof of Concept and
10 Mechanism Study.” *Molecular Psychiatry* 28 (2): 564–78.
11 <https://doi.org/10.1038/s41380-022-01836-9>.

12 Augusto, Danillo G., Lawton D. Murdolo, Demetra S. M. Chatzileontiadou, Joseph J.
13 Sabatino, Tasneem Yusufali, Noah D. Peyser, Xochitl Butcher, et al. 2023. “A
14 Common Allele of HLA Is Associated with Asymptomatic SARS-CoV-2 Infection.”
15 *Nature* 620 (7972): 128–36. <https://doi.org/10.1038/s41586-023-06331-x>.

16 Backman, Joshua D., Alexander H. Li, Anthony Marcketta, Dylan Sun, Joelle Mbatchou,
17 Michael D. Kessler, Christian Benner, et al. 2021. “Exome Sequencing and
18 Analysis of 454,787 UK Biobank Participants.” *Nature* 599 (7886): 628–34.
19 <https://doi.org/10.1038/s41586-021-04103-z>.

20 Bowden, Jack, George Davey Smith, and Stephen Burgess. 2015. “Mendelian
21 Randomization with Invalid Instruments: Effect Estimation and Bias Detection
22 through Egger Regression.” *International Journal of Epidemiology* 44 (2):
23 512–25. <https://doi.org/10.1093/ije/dyv080>.

- 1 Bowden, Jack, George Davey Smith, Philip C. Haycock, and Stephen Burgess. 2016.
2 “Consistent Estimation in Mendelian Randomization with Some Invalid
3 Instruments Using a Weighted Median Estimator.” *Genetic Epidemiology* 40 (4):
4 304–14. <https://doi.org/10.1002/gepi.21965>.
- 5 Breiman, Adrien, Nathalie Ruvèn-Clouet, and Jacques Le Pendu. 2020. “Harnessing the
6 Natural Anti-Glycan Immune Response to Limit the Transmission of Enveloped
7 Viruses Such as SARS-CoV-2.” *PLoS Pathogens* 16 (5): e1008556.
8 <https://doi.org/10.1371/journal.ppat.1008556>.
- 9 Brenu, Ekuwa W., Mieke L. van Driel, Donald R. Staines, Kevin J. Ashton, Sharni L.
10 Hardcastle, James Keane, Lotti Tajouri, Daniel Peterson, Sandra B. Ramos, and
11 Sonya M. Marshall-Gradisnik. 2012. “Longitudinal Investigation of Natural Killer
12 Cells and Cytokines in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis.”
13 *Journal of Translational Medicine* 10 (May):88.
14 <https://doi.org/10.1186/1479-5876-10-88>.
- 15 Bugert, Peter, Gabriele Rink, Katharina Kemp, and Harald Klüter. 2012. “Blood Group
16 ABO Genotyping in Paternity Testing.” *Transfusion Medicine and Hemotherapy:*
17 *Offizielles Organ Der Deutschen Gesellschaft Fur Transfusionsmedizin Und*
18 *Immunhamatologie* 39 (3): 182–86. <https://doi.org/10.1159/000339235>.
- 19 Bulik-Sullivan, Brendan K., Po-Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang,
20 Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick
21 Patterson, Mark J. Daly, Alkes L. Price, and Benjamin M. Neale. 2015. “LD Score
22 Regression Distinguishes Confounding from Polygenicity in Genome-Wide
23 Association Studies.” *Nature Genetics* 47 (3): 291–95.

- 1 <https://doi.org/10.1038/ng.3211>.
- 2 Bull-Otterson, Lara, Sarah Baca, Sharon Saydah, Tegan K. Boehmer, Stacey Adjei,
3 Simone Gray, and Aaron M. Harris. 2022. "Post-COVID Conditions Among Adult
4 COVID-19 Survivors Aged 18–64 and ≥65 Years — United States, March
5 2020–November 2021." *MMWR. Morbidity and Mortality Weekly Report* 71 (21):
6 713–17. <https://doi.org/10.15585/mmwr.mm7121e1>.
- 7 Català, Martí, Núria Mercadé-Besora, Raivo Kolde, Nhung T H Trinh, Elena Roel,
8 Edward Burn, Trishna Rathod-Mistry, et al. 2024. "The Effectiveness of
9 COVID-19 Vaccines to Prevent Long COVID Symptoms: Staggered Cohort Study
10 of Data from the UK, Spain, and Estonia." *The Lancet Respiratory Medicine* 12
11 (3): 225–36. [https://doi.org/10.1016/S2213-2600\(23\)00414-9](https://doi.org/10.1016/S2213-2600(23)00414-9).
- 12 Cervia-Hasler, Carlo, Sarah C. Brüningk, Tobias Hoch, Bowen Fan, Giulia Muzio, Ryan
13 C. Thompson, Laura Ceglarek, et al. 2024. "Persistent Complement
14 Dysregulation with Signs of Thromboinflammation in Active Long Covid." *Science*
15 (*New York, N.Y.*) 383 (6680): eadg7942.
16 <https://doi.org/10.1126/science.adg7942>.
- 17 Chaudhary, Ninad S, Catherine H Weldon, Priyanka Nandakumar, Janie Shelton,
18 23andMe Research Team, Michael V Holmes, and Stella Aslibekyan. 2024.
19 "Multi-Ancestry GWAS of Diarrhea during Acute SARS-CoV2 Infection Identifies
20 Multiple Novel Loci and Contrasting Etiological Roles of Irritable Bowel Syndrome
21 Subtypes." <https://doi.org/10.1101/2024.04.03.24305274>.
- 22 Cohen, Ken, Sheng Ren, Kevin Heath, Micah C. Dasmariñas, Karol Giuseppe Jubilo,
23 Yinglong Guo, Marc Lipsitch, and Sarah E. Daugherty. 2022. "Risk of Persistent

- 1 and New Clinical Sequelae among Adults Aged 65 Years and Older during the
2 Post-Acute Phase of SARS-CoV-2 Infection: Retrospective Cohort Study.” *BMJ*
3 (*Clinical Research Ed.*) 376 (February):e068414.
4 <https://doi.org/10.1136/bmj-2021-068414>.
- 5 COVID-19 Host Genetics Initiative. 2023. “A Second Update on Mapping the Human
6 Genetic Architecture of COVID-19.” *Nature* 621 (7977): E7–26.
7 <https://doi.org/10.1038/s41586-023-06355-3>.
- 8 Craig, David W., Elizabeth Hutchins, Ivo Violich, Eric Alsop, J. Raphael Gibbs, Shawn
9 Levy, Madison Robison, et al. 2021. “RNA Sequencing of Whole Blood Reveals
10 Early Alterations in Immune Cells and Gene Expression in Parkinson’s Disease.”
11 *Nature Aging* 1 (8): 734–47. <https://doi.org/10.1038/s43587-021-00088-6>.
- 12 Davis, Hannah E., Gina S. Assaf, Lisa McCorkell, Hannah Wei, Ryan J. Low, Yochai
13 Re’em, Signe Redfield, Jared P. Austin, and Athena Akrami. 2021.
14 “Characterizing Long COVID in an International Cohort: 7 Months of Symptoms
15 and Their Impact.” *eClinicalMedicine* 38 (August):101019.
16 <https://doi.org/10.1016/j.eclinm.2021.101019>.
- 17 Davis, Hannah E., Lisa McCorkell, Julia Moore Vogel, and Eric J. Topol. 2023. “Long
18 COVID: Major Findings, Mechanisms and Recommendations.” *Nature Reviews*
19 *Microbiology* 21 (3): 133–46. <https://doi.org/10.1038/s41579-022-00846-2>.
- 20 Debebe, Bisrat J, Lies Boelen, James C Lee, IAVI Protocol C Investigators, Eduard J
21 Sanders, Omu Anzala, Anatoli Kamali, et al. 2020. “Identifying the Immune
22 Interactions Underlying HLA Class I Disease Associations.” *eLife* 9
23 (April):e54558. <https://doi.org/10.7554/eLife.54558>.

- 1 Drevets, Wayne C., Gayle M. Wittenberg, Edward T. Bullmore, and Husseini K. Manji.
2 2022. “Immune Targets for Therapeutic Development in Depression: Towards
3 Precision Medicine.” *Nature Reviews Drug Discovery* 21 (3): 224–44.
4 <https://doi.org/10.1038/s41573-021-00368-1>.
- 5 Durand, Eric Y, Chuong B Do, Joanna L Mountain, and J. Michael Macpherson. 2014.
6 “Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution.”
7 Preprint. Bioinformatics. <https://doi.org/10.1101/010512>.
- 8 Ferkingstad, Egil, Patrick Sulem, Bjarni A. Atlason, Gardar Sveinbjornsson, Magnus I.
9 Magnusson, Edda L. Styrismisdottir, Kristbjorg Gunnarsdottir, et al. 2021.
10 “Large-Scale Integration of the Plasma Proteome with Genetics and Disease.”
11 *Nature Genetics* 53 (12): 1712–21. <https://doi.org/10.1038/s41588-021-00978-w>.
- 12 Ford, Nicole D., Abraham Agedew, Alexandra F. Dalton, Jordan Singleton, Cria G.
13 Perrine, and Sharon Saydah. 2024. “Notes from the Field: Long COVID
14 Prevalence Among Adults — United States, 2022.” *MMWR. Morbidity and
15 Mortality Weekly Report* 73 (6): 135–36.
16 <https://doi.org/10.15585/mmwr.mm7306a4>.
- 17 Franzén, Oscar, Raili Ermel, Ariella Cohain, Nicholas K. Akers, Antonio Di Narzo,
18 Husain A. Talukdar, Hassan Foroughi-Asl, et al. 2016. “Cardiometabolic Risk Loci
19 Share Downstream Cis- and Trans-Gene Regulation across Tissues and
20 Diseases.” *Science (New York, N.Y.)* 353 (6301): 827–30.
21 <https://doi.org/10.1126/science.aad6970>.
- 22 Griffith, Gareth J., Tim T. Morris, Matthew J. Tudball, Annie Herbert, Giulia Mancano,
23 Lindsey Pike, Gemma C. Sharp, et al. 2020. “Collider Bias Undermines Our

- 1 Understanding of COVID-19 Disease Risk and Severity.” *Nature Communications*
2 11 (1): 5749. <https://doi.org/10.1038/s41467-020-19478-2>.
- 3 GTEEx Consortium. 2020. “The GTEEx Consortium Atlas of Genetic Regulatory Effects
4 across Human Tissues.” *Science (New York, N.Y.)* 369 (6509): 1318–30.
5 <https://doi.org/10.1126/science.aaz1776>.
- 6 Guillon, Patrice, Monique Clément, Véronique Sébille, Jean-Gérard Rivain, Chih-Fong
7 Chou, Nathalie Ruvoën-Clouet, and Jacques Le Pendu. 2008. “Inhibition of the
8 Interaction between the SARS-CoV Spike Protein and Its Cellular Receptor by
9 Anti-Histo-Blood Group Antibodies.” *Glycobiology* 18 (12): 1085–93.
10 <https://doi.org/10.1093/glycob/cwn093>.
- 11 Hajdarevic, Riad, Asgeir Lande, Jesper Mehlsen, Anne Rydland, Daisy D. Sosa, Elin B.
12 Strand, Olav Mella, et al. 2022. “Genetic Association Study in Myalgic
13 Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) Identifies Several
14 Potential Risk Loci.” *Brain, Behavior, and Immunity* 102 (May):362–69.
15 <https://doi.org/10.1016/j.bbi.2022.03.010>.
- 16 Harrow, Jennifer, Adam Frankish, Jose M. Gonzalez, Electra Tapanari, Mark Diekhans,
17 Felix Kokocinski, Bronwen L. Aken, et al. 2012. “GENCODE: The Reference
18 Human Genome Annotation for The ENCODE Project.” *Genome Research* 22
19 (9): 1760–74. <https://doi.org/10.1101/gr.135350.111>.
- 20 Hemani, Gibran, Kate Tilling, and George Davey Smith. 2017. “Orienting the Causal
21 Relationship between Imprecisely Measured Traits Using GWAS Summary
22 Data.” *PLoS Genetics* 13 (11): e1007081.
23 <https://doi.org/10.1371/journal.pgen.1007081>.

- 1 Howard, David M., Mark J. Adams, Toni-Kim Clarke, Jonathan D. Hafferty, Jude Gibson,
2 Masoud Shirali, Jonathan R. I. Coleman, et al. 2019. "Genome-Wide
3 Meta-Analysis of Depression Identifies 102 Independent Variants and Highlights
4 the Importance of the Prefrontal Brain Regions." *Nature Neuroscience* 22 (3):
5 343–52. <https://doi.org/10.1038/s41593-018-0326-7>.
- 6 Howard, David M., Mark J. Adams, Masoud Shirali, Toni-Kim Clarke, Riccardo E.
7 Marioni, Gail Davies, Jonathan R. I. Coleman, et al. 2018. "Genome-Wide
8 Association Study of Depression Phenotypes in UK Biobank Identifies Variants in
9 Excitatory Synaptic Pathways." *Nature Communications* 9 (1): 1470.
10 <https://doi.org/10.1038/s41467-018-03819-3>.
- 11 International HIV Controllers Study, Florencia Pereyra, Xiaoming Jia, Paul J. McLaren,
12 Amalio Telenti, Paul I. W. de Bakker, Bruce D. Walker, et al. 2010. "The Major
13 Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation."
14 *Science (New York, N.Y.)* 330 (6010): 1551–57.
15 <https://doi.org/10.1126/science.1195271>.
- 16 Kandikattu, Hemanth Kumar, Sathisha Upparahalli Venkateshaiah, Sandeep Kumar,
17 and Anil Mishra. 2020. "IL-15 Immunotherapy Is a Viable Strategy for COVID-19."
18 *Cytokine & Growth Factor Reviews* 54 (August):24–31.
19 <https://doi.org/10.1016/j.cytogfr.2020.06.008>.
- 20 Kerimov, Nurlan, James D. Hayhurst, Kateryna Peikova, Jonathan R. Manning, Peter
21 Walter, Liis Kolberg, Marija Samoviča, et al. 2021. "A Compendium of Uniformly
22 Processed Human Gene Expression and Splicing Quantitative Trait Loci." *Nature*
23 *Genetics* 53 (9): 1290–99. <https://doi.org/10.1038/s41588-021-00924-w>.

- 1 Kettunen, Johannes, Taru Tukiainen, Antti-Pekka Sarin, Alfredo Ortega-Alonso, Emmi
2 Tikkanen, Leo-Pekka Lyytikäinen, Antti J. Kangas, et al. 2012. “Genome-Wide
3 Association Study Identifies Multiple Loci Influencing Human Serum Metabolite
4 Levels.” *Nature Genetics* 44 (3): 269–76. <https://doi.org/10.1038/ng.1073>.
- 5 Klein, Jon, Jamie Wood, Jillian R. Jaycox, Rahul M. Dhodapkar, Peiwen Lu, Jeff R.
6 Gehlhausen, Alexandra Tabachnikova, et al. 2023. “Distinguishing Features of
7 Long COVID Identified through Immune Profiling.” *Nature* 623 (7985): 139–48.
8 <https://doi.org/10.1038/s41586-023-06651-y>.
- 9 Knuppel, Anika, Andy Boyd, John Macleod, Nishi Chaturvedi, and Dylan M Williams.
10 2023. “The Long COVID Evidence Gap: Comparing Self-Reporting and Clinical
11 Coding of Long COVID Using Longitudinal Study Data Linked to Healthcare
12 Records.” <https://doi.org/10.1101/2023.02.10.23285717>.
- 13 Koolpe, G. A., W. L. Nelson, T. L. Gioannini, L. Angel, N. Appelmans, and E. J. Simon.
14 1985. “Opioid Agonists and Antagonists. 6-Desoxy-6-Substituted Lactone,
15 Epoxide, and Glycidate Ester Derivatives of Naltrexone and Oxymorphone.”
16 *Journal of Medicinal Chemistry* 28 (7): 949–57.
17 <https://doi.org/10.1021/jm00145a018>.
- 18 Lammi, Vilma, Tomoko Nakanishi, Samuel E. Jones, Shea J. Andrews, Juha
19 Karjalainen, Beatriz Cortés, Heath E. O’Brien, et al. 2023. “Genome-Wide
20 Association Study of Long COVID.” Preprint. *Genetic and Genomic Medicine*.
21 <https://doi.org/10.1101/2023.06.29.23292056>.
- 22 Lappalainen, Tuuli, Michael Sammeth, Marc R. Friedländer, Peter A. C. ’t Hoen, Jean
23 Monlong, Manuel A. Rivas, Mar González-Porta, et al. 2013. “Transcriptome and

- 1 Genome Sequencing Uncovers Functional Variation in Humans.” *Nature* 501
2 (7468): 506–11. <https://doi.org/10.1038/nature12531>.
- 3 Ledford, Heidi. 2024. “Long COVID Is a Double Curse in Low-Income Nations — Here’s
4 Why.” *Nature* 625 (7993): 20–22. <https://doi.org/10.1038/d41586-023-04088-x>.
- 5 Lopez-Leon, Sandra, Talia Wegman-Ostrosky, Carol Perelman, Rosalinda Sepulveda,
6 Paulina A. Rebolledo, Angelica Cuapio, and Sonia Villapol. 2021. “More than 50
7 Long-Term Effects of COVID-19: A Systematic Review and Meta-Analysis.”
8 *Scientific Reports* 11 (1): 16144. <https://doi.org/10.1038/s41598-021-95565-8>.
- 9 Lu, Ting, Rui Ma, Wenjuan Dong, Kun-Yu Teng, Daniel S. Kollath, Zhiyao Li, Jinhee Yi,
10 et al. 2022. “Off-the-Shelf CAR Natural Killer Cells Secreting IL-15 Target Spike
11 in Treating COVID-19.” *Nature Communications* 13 (1): 2576.
12 <https://doi.org/10.1038/s41467-022-30216-8>.
- 13 Miyamoto, Yoichi, Yumi Itoh, Tatsuya Suzuki, Tomohisa Tanaka, Yusuke Sakai, Masaru
14 Koido, Chiaki Hata, et al. 2021. “SARS-CoV-2 ORF6 Disturbs Nucleocytoplasmic
15 Trafficking to Advance the Viral Replication.”
16 <https://doi.org/10.1101/2021.02.24.432656>.
- 17 Naik, Hiten, Karen C. Tran, John A. Staples, Roy H. Perlis, and Adeera Levin. 2024.
18 “Psychiatric Symptoms, Treatment Uptake, and Barriers to Mental Health Care
19 Among US Adults With Post-COVID-19 Condition.” *JAMA Network Open* 7 (4):
20 e248481. <https://doi.org/10.1001/jamanetworkopen.2024.8481>.
- 21 National Library of Medicine (U.S.). 2020. “A Phase I/II Study of Universal Off-the-Shelf
22 NKG2D-ACE2 CAR-NK Cells for Therapy of COVID-19. Identifier
23 NCT04324996.” <https://clinicaltrials.gov/ct2/show/NCT04324996>.

- 1 Norland, Kristjan, Gardar Sveinbjornsson, Rosa B. Thorolfsdottir, Olafur B. Davidsson,
2 Vinicius Tragante, Sridharan Rajamani, Anna Helgadottir, et al. 2019. "Sequence
3 Variants with Large Effects on Cardiac Electrophysiology and Disease." *Nature*
4 *Communications* 10 (1): 4803. <https://doi.org/10.1038/s41467-019-12682-9>.
- 5 Notarte, Kin Israel, Jesus Alfonso Catahay, Jacqueline Veronica Velasco, Adriel
6 Pastrana, Abbygail Therese Ver, Flos Carmeli Pangilinan, Princess Juneire
7 Peligro, et al. 2022. "Impact of COVID-19 Vaccination on the Risk of Developing
8 Long-COVID and on Existing Long-COVID Symptoms: A Systematic Review."
9 *eClinicalMedicine* 53 (November):101624.
10 <https://doi.org/10.1016/j.eclinm.2022.101624>.
- 11 O'Mahoney, Lauren L., Ash Routen, Clare Gillies, Winifred Ekezie, Anneka Welford,
12 Alexa Zhang, Urvi Karamchandani, et al. 2023. "The Prevalence and Long-Term
13 Health Effects of Long Covid among Hospitalised and Non-Hospitalised
14 Populations: A Systematic Review and Meta-Analysis." *eClinicalMedicine* 55
15 (January):101762. <https://doi.org/10.1016/j.eclinm.2022.101762>.
- 16 Ongen, Halit, Alfonso Buil, Andrew Anand Brown, Emmanouil T. Dermitzakis, and
17 Olivier Delaneau. 2016. "Fast and Efficient QTL Mapper for Thousands of
18 Molecular Phenotypes." *Bioinformatics* 32 (10): 1479–85.
19 <https://doi.org/10.1093/bioinformatics/btv722>.
- 20 Pairo-Castineira, Erola, Konrad Rawlik, Andrew D. Bretherick, Ting Qi, Yang Wu, Isar
21 Nassiri, Glenn A. McConkey, et al. 2023. "GWAS and Meta-Analysis Identifies 49
22 Genetic Variants Underlying Critical COVID-19." *Nature* 617 (7962): 764–68.
23 <https://doi.org/10.1038/s41586-023-06034-3>.

- 1 Peter, Raphael S, Alexandra Nieters, Hans-Georg Kräusslich, Stefan O Brockmann, Siri
2 Göpel, Gerhard Kindle, Uta Merle, Jürgen M Steinacker, Dietrich Rothenbacher,
3 and Winfried V Kern. 2022. “Post-Acute Sequelae of Covid-19 Six to 12 Months
4 after Infection: Population Based Study.” *BMJ*, October, e071050.
5 <https://doi.org/10.1136/bmj-2022-071050>.
- 6 Qiu, Chengxiang, Shizheng Huang, Jihwan Park, YoSon Park, Yi-An Ko, Matthew J.
7 Seasock, Joshua S. Bryer, et al. 2018. “Renal Compartment-Specific Genetic
8 Variation Analyses Identify New Pathways in Chronic Kidney Disease.” *Nature*
9 *Medicine* 24 (11): 1721–31. <https://doi.org/10.1038/s41591-018-0194-4>.
- 10 Raj, Towfique, Katie Rothamel, Sara Mostafavi, Chun Ye, Mark N. Lee, Joseph M.
11 Replogle, Ting Feng, et al. 2014. “Polarization of the Effects of Autoimmune and
12 Neurodegenerative Risk Alleles in Leukocytes.” *Science (New York, N. Y.)* 344
13 (6183): 519–23. <https://doi.org/10.1126/science.1249547>.
- 14 Rivas, Jose Luis, Teresa Palencia, Guerau Fernández, and Milagros García. 2018.
15 “Association of T and NK Cell Phenotype With the Diagnosis of Myalgic
16 Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS).” *Frontiers in*
17 *Immunology* 9 (May):1028. <https://doi.org/10.3389/fimmu.2018.01028>.
- 18 Robertson, McKaylee M, Saba A Qasmieh, Sarah G Kulkarni, Chloe A Teasdale, Heidi
19 E Jones, Margaret McNairy, Luisa N Borrell, and Denis Nash. 2023. “The
20 Epidemiology of Long Coronavirus Disease in US Adults.” *Clinical Infectious*
21 *Diseases* 76 (9): 1636–45. <https://doi.org/10.1093/cid/ciac961>.
- 22 Severe Covid-19 GWAS Group, David Ellinghaus, Frauke Degenhardt, Luis Bujanda,
23 Maria Buti, Agustín Albillos, Pietro Invernizzi, et al. 2020. “Genomewide

- 1 Association Study of Severe Covid-19 with Respiratory Failure.” *The New*
2 *England Journal of Medicine* 383 (16): 1522–34.
3 <https://doi.org/10.1056/NEJMoa2020283>.
- 4 Shankar, Vishnu, Julie Wilhelmy, Ellis J. Curtis, Basil Michael, Layla Cervantes, Vamsee
5 A. Mallajosyula, Ronald W. Davis, et al. 2024. “Oxidative Stress Is a Shared
6 Characteristic of ME/CFS and Long COVID.”
7 <https://doi.org/10.1101/2024.05.04.592477>.
- 8 Shelton, Janie F., Anjali J. Shastri, Chelsea Ye, Catherine H. Weldon, Teresa
9 Filshtein-Sonmez, Daniella Coker, Antony Symons, et al. 2021. “Trans-Ancestry
10 Analysis Reveals Genetic and Nongenetic Associations with COVID-19
11 Susceptibility and Severity.” *Nature Genetics* 53 (6): 801–8.
12 <https://doi.org/10.1038/s41588-021-00854-7>.
- 13 Sherif, Zaki A., Christian R. Gomez, Thomas J. Connors, Timothy J. Henrich, William
14 Brian Reeves, and RECOVER Mechanistic Pathway Task Force. 2023.
15 “Pathogenic Mechanisms of Post-Acute Sequelae of SARS-CoV-2 Infection
16 (PASC).” *eLife* 12 (March):e86002. <https://doi.org/10.7554/eLife.86002>.
- 17 Soriano, Joan B., Srinivas Murthy, John C. Marshall, Pryanka Relan, Janet V. Diaz, and
18 WHO Clinical Case Definition Working Group on Post-COVID-19 Condition.
19 2022. “A Clinical Case Definition of Post-COVID-19 Condition by a Delphi
20 Consensus.” *The Lancet. Infectious Diseases* 22 (4): e102–7.
21 [https://doi.org/10.1016/S1473-3099\(21\)00703-9](https://doi.org/10.1016/S1473-3099(21)00703-9).
- 22 Steffen, Brian T, James S Pankow, Pamela L Lutsey, Ryan T Demmer, Jeffrey R
23 Misialek, Weihua Guan, Logan T Cowan, Josef Coresh, Faye L Norby, and

- 1 Weihong Tang. 2022. “Proteomic Profiling Identifies Novel Proteins for Genetic
2 Risk of Severe COVID-19: The Atherosclerosis Risk in Communities Study.”
3 *Human Molecular Genetics* 31 (14): 2452–61.
4 <https://doi.org/10.1093/hmg/ddac024>.
- 5 Stegle, Oliver, Leopold Parts, Matias Piipari, John Winn, and Richard Durbin. 2012.
6 “Using Probabilistic Estimation of Expression Residuals (PEER) to Obtain
7 Increased Power and Interpretability of Gene Expression Analyses.” *Nature*
8 *Protocols* 7 (3): 500–507. <https://doi.org/10.1038/nprot.2011.457>.
- 9 Storey, John D. 2002. “A Direct Approach to False Discovery Rates.” *Journal of the*
10 *Royal Statistical Society Series B: Statistical Methodology* 64 (3): 479–98.
11 <https://doi.org/10.1111/1467-9868.00346>.
- 12 Su, Sizhen, Yimiao Zhao, Na Zeng, Xiaoxing Liu, Yongbo Zheng, Jie Sun, Yi Zhong, et
13 al. 2023. “Epidemiology, Clinical Presentation, Pathophysiology, and
14 Management of Long COVID: An Update.” *Molecular Psychiatry*, July.
15 <https://doi.org/10.1038/s41380-023-02171-3>.
- 16 Taquet, Maxime, Quentin Dercon, Sierra Luciano, John R. Geddes, Masud Husain, and
17 Paul J. Harrison. 2021. “Incidence, Co-Occurrence, and Evolution of
18 Long-COVID Features: A 6-Month Retrospective Cohort Study of 273,618
19 Survivors of COVID-19.” *PLoS Medicine* 18 (9): e1003773.
20 <https://doi.org/10.1371/journal.pmed.1003773>.
- 21 Taylor, Krystyna, Matthew Pearson, Sayoni Das, Jason Sardell, Karolina Chocian, and
22 Steve Gardner. 2023. “Genetic Risk Factors for Severe and Fatigue Dominant
23 Long COVID and Commonalities with ME/CFS Identified by Combinatorial

- 1 Analysis.” *Journal of Translational Medicine* 21 (1): 775.
- 2 <https://doi.org/10.1186/s12967-023-04588-4>.
- 3 Thompson, Ryan C., Nicole W. Simons, Lillian Wilkins, Esther Cheng, Diane Marie Del
4 Valle, Gabriel E. Hoffman, Carlo Cervia, et al. 2023. “Molecular States during
5 Acute COVID-19 Reveal Distinct Etiologies of Long-Term Sequelae.” *Nature*
6 *Medicine* 29 (1): 236–46. <https://doi.org/10.1038/s41591-022-02107-4>.
- 7 Tian, Chao, Bethann S. Hromatka, Amy K. Kiefer, Nicholas Eriksson, Suzanne M.
8 Noble, Joyce Y. Tung, and David A. Hinds. 2017. “Genome-Wide Association and
9 HLA Region Fine-Mapping Studies Identify Susceptibility Loci for Multiple
10 Common Infections.” *Nature Communications* 8 (1): 599.
11 <https://doi.org/10.1038/s41467-017-00257-5>.
- 12 Tran, Viet-Thi, Elodie Perrodeau, Julia Saldanha, Isabelle Pane, and Philippe Ravaud.
13 2023. “Efficacy of First Dose of Covid-19 Vaccine versus No Vaccination on
14 Symptoms of Patients with Long Covid: Target Trial Emulation Based on
15 ComPaRe e-Cohort.” *BMJ Medicine* 2 (1): e000229.
16 <https://doi.org/10.1136/bmjmed-2022-000229>.
- 17 Tsampasian, Vasiliki, Hussein Elghazaly, Rahul Chattopadhyay, Maciej Debski, Thin Kyi
18 Phyu Naing, Pankaj Garg, Allan Clark, Eleana Ntatsaki, and Vassilios S.
19 Vassiliou. 2023. “Risk Factors Associated With Post-COVID-19 Condition: A
20 Systematic Review and Meta-Analysis.” *JAMA Internal Medicine* 183 (6):
21 566–80. <https://doi.org/10.1001/jamainternmed.2023.0750>.
- 22 Ursini, Francesco, Jacopo Ciaffi, Luana Mancarella, Lucia Lisi, Veronica Brusi, Carlotta
23 Cavallari, Martina D’Onghia, et al. 2021. “Fibromyalgia: A New Facet of the

- 1 Post-COVID-19 Syndrome Spectrum? Results from a Web-Based Survey.” *RMD*
2 *Open* 7 (3): e001735. <https://doi.org/10.1136/rmdopen-2021-001735>.
- 3 Vasan, Senthil K., Klaus Rostgaard, Ammar Majeed, Henrik Ullum, Kjell-Einar Titlestad,
4 Ole B. V. Pedersen, Christian Erikstrup, et al. 2016. “ABO Blood Group and Risk
5 of Thromboembolic and Arterial Disease: A Study of 1.5 Million Blood Donors.”
6 *Circulation* 133 (15): 1449–57; discussion 1457.
7 <https://doi.org/10.1161/CIRCULATIONAHA.115.017563>.
- 8 Verma, Vivek, Gillian L. Drury, Marc Parisien, Ayşe N. Özdağ Acarli, Tho-Alfakar
9 Al-Aubodah, Anastasia Nijnik, Xia Wen, et al. 2022. “Unbiased Immune Profiling
10 Reveals a Natural Killer Cell-Peripheral Nerve Axis in Fibromyalgia.” *Pain* 163
11 (7): e821–36. <https://doi.org/10.1097/j.pain.0000000000002498>.
- 12 Walther, Andreas, Anne Mackens-Kiani, Julian Eder, Maik Herbig, Christoph Herold,
13 Clemens Kirschbaum, Jochen Guck, Lucas Daniel Wittwer, Katja Beesdo-Baum,
14 and Martin Kräter. 2022. “Depressive Disorders Are Associated with Increased
15 Peripheral Blood Cell Deformability: A Cross-Sectional Case-Control Study
16 (Mood-Morph).” *Translational Psychiatry* 12 (1): 150.
17 <https://doi.org/10.1038/s41398-022-01911-3>.
- 18 Wang, Gao, Abhishek Sarkar, Peter Carbonetto, and Matthew Stephens. 2020. “A
19 Simple New Approach to Variable Selection in Regression, with Application to
20 Genetic Fine Mapping.” *Journal of the Royal Statistical Society Series B:*
21 *Statistical Methodology* 82 (5): 1273–1300. <https://doi.org/10.1111/rssb.12388>.
- 22 Wellcome Trust Case Control Consortium, Julian B. Maller, Gilean McVean, Jake
23 Byrnes, Damjan Vukcevic, Kimmo Palin, Zhan Su, et al. 2012. “Bayesian

- 1 Refinement of Association Signals for 14 Loci in 3 Common Diseases.” *Nature*
2 *Genetics* 44 (12): 1294–1301. <https://doi.org/10.1038/ng.2435>.
- 3 Whitaker, Matthew, Joshua Elliott, Marc Chadeau-Hyam, Steven Riley, Ara Darzi,
4 Graham Cooke, Helen Ward, and Paul Elliott. 2022. “Persistent COVID-19
5 Symptoms in a Community Study of 606,434 People in England.” *Nature*
6 *Communications* 13 (1): 1957. <https://doi.org/10.1038/s41467-022-29521-z>.
- 7 Wu, Peng, Lin Ding, Xiaodong Li, Siyang Liu, Fanjun Cheng, Qing He, Mingzhong Xiao,
8 et al. 2021. “Trans-Ethnic Genome-Wide Association Study of Severe
9 COVID-19.” *Communications Biology* 4 (1): 1034.
10 <https://doi.org/10.1038/s42003-021-02549-5>.
- 11 Yuen, Chun-Kit, Joy-Yan Lam, Wan-Man Wong, Long-Fung Mak, Xiaohui Wang, Hin
12 Chu, Jian-Piao Cai, et al. 2020. “SARS-CoV-2 Nsp13, Nsp14, Nsp15 and Orf6
13 Function as Potent Interferon Antagonists.” *Emerging Microbes & Infections* 9
14 (1): 1418–28. <https://doi.org/10.1080/22221751.2020.1780953>.
- 15 Zhang, Jingning, Diptavo Dutta, Anna Köttgen, Adrienne Tin, Pascal Schlosser, Morgan
16 E. Grams, Benjamin Harvey, et al. 2022. “Plasma Proteome Analyses in
17 Individuals of European and African Ancestry Identify Cis-pQTLs and Models for
18 Proteome-Wide Association Studies.” *Nature Genetics* 54 (5): 593–602.
19 <https://doi.org/10.1038/s41588-022-01051-w>.
- 20 Zhao, Jiao, Yan Yang, Hanping Huang, Dong Li, Dongfeng Gu, Xiangfeng Lu, Zheng
21 Zhang, et al. 2021. “Relationship Between the ABO Blood Group and the
22 Coronavirus Disease 2019 (COVID-19) Susceptibility.” *Clinical Infectious*
23 *Diseases: An Official Publication of the Infectious Diseases Society of America*

1 73 (2): 328–31. <https://doi.org/10.1093/cid/ciaa1150>.

2 Zheng, X., J. Shen, C. Cox, J. C. Wakefield, M. G. Ehm, M. R. Nelson, and B. S. Weir.
3 2014. “HIBAG--HLA Genotype Imputation with Attribute Bagging.” *The*
4 *Pharmacogenomics Journal* 14 (2): 192–200. <https://doi.org/10.1038/tpj.2013.18>.

5 Zorina-Lichtenwalter, Katerina, Carmen I. Bango, Lukas Van Oudenhove, Marta Čeko,
6 Martin A. Lindquist, Andrew D. Grotzinger, Matthew C. Keller, Naomi P.
7 Friedman, and Tor D. Wager. 2023. “Genetic Risk Shared across 24 Chronic Pain
8 Conditions: Identification and Characterization with Genomic Structural Equation
9 Modeling.” *Pain* 164 (10): 2239–52.
10 <https://doi.org/10.1097/j.pain.0000000000002922>.

11

12