Study Protocol

|  |  |
| --- | --- |
| STUDY FULL TITLE | Using whole-genome or whole-exome sequencing to identify host genetic determinants of Covid-19 susceptibility and severity |
| STUDY PROTOCOL VERSION | 4.0 |
| STUDY PROTOCOL VERSION DATE | 2020-11-17 |
| Cohorts (alphabetical order) | **Biobanque Québec Covid-19**  Investigator: J Brent Richards, Professor of Medicine, McGill University  Co-Investigators: Guillaume Butler-Laporte, Tomoko Nakanishi, Sirui Zhou, Vincenzo Forgetta  **CHIRP Study**  Investigator: Sulggi Lee  **Columbia COVID-19 Biobank**  Investigators: Krzysztof Kiryluk, David Goldstein  Co-investigator: Gundula Povysil  **Deutsche COVID-19 OMICS Initiative (DeCOI)**  Investigators: Olaf Riess, Eva Schulte, Julien Gagneur  **GEN-COVID**  Investigator: Alessandra Renieri  Co-investigator: Simone Furini  **Helix and Healthy Nevada Project Exome+ Covid-19 Phenotypes**  Investigator: Liz Cirulli  **INMUGEN-CoV2**  Investigators: Anna Planas, Israel Fernández-Cadenas, Jordi Pérez-Tur  **POLCOVID-Genomika**  Investigator: Miroslaw Kwasniewski  Co-investigator: Pawel Olszewski  **Qatar Genome**  Investigator: Sail Ismail  Co-investigator: Hamdi Mbarek  **Regeneron Pharmaceuticals Inc.**  Investigator: Manuel Allen Revez Ferreira  Co-investigator: Jack Kosmicki  **Saudi Genome Program**  Investigator: Malak Abedalthagafi  Co-investigator: Manal Alaamery  **Sinai\_Covid**  Investigator: Joseph Buxbaum  **Swedish Covid-19 Cohort**  Investigator: Hugo Zeberg  **UC COVID-19 Host Genetics Consortium**  Investigators: Dan Geschwind, Manish Butte, Bogdan Pasaniuc, Jonathan Sebat |

# Table of Contents

[1](#_1fob9te) Table of Contents 3

[2](#_2et92p0) Abbreviations 4

[3](#_1t3h5sf) Introduction 4

[3.1](#_4d34og8) Preface 4

[3.2](#_2s8eyo1) Scope of the analyses 4

[4](#_17dp8vu) Study Objectives and Endpoints 5

[4.1](#_3rdcrjn) Study Objectives 5

[4.2](#_lnxbz9) Endpoints, Outcomes, and Measurements 5

[4.3](#_35nkun2) Genetic loci 6

[5](#_1ksv4uv) Study Methods 6

[5.1](#_44sinio) Participants inclusion criteria 6

[5.2](#_2jxsxqh) Whole genome or Exome sequencing 6

[5.3](#_1y810tw) Genetic variant calling and quality control 6

[5.4](#_4i7ojhp) Genetic variant annotation 6

[5.5](#_2xcytpi) Genetic association studies 7

[5.6](#_1ci93xb) Meta-analysis 8

[6](#_2bn6wsx) Data sharing 9

[7](#_3as4poj) References 10

[8](#_1pxezwc) Appendix 1 11

[9](#_2p2csry) Appendix 2 19

# 

# Abbreviations

|  |  |
| --- | --- |
| BiPAP | Bilevel positive airway pressure |
| CPAP | Continuous positive airway pressure |
| CK | Creatine kinase |
| Covid-19 HGI | Covid-19 Host Genetic Initiative |
| CRP | C-reactive protein |
| ECMO | Extracorporeal membrane oxygenation |
| GWAS | Genome-wide association study |
| LDH | Lactate dehydrogenase |
| LoF | Loss of function genetic variants |
| GWAS | Genome-wide association study |
| NAAT | Nucleic acid amplification test |
| PT | Prothrombin time |
| SNP | Single nucleotide polymorphism |
| WES | Whole-exome sequencing |
| WGS | Whole-genome sequencing |

# 

# Introduction

## Preface

The Covid-19 pandemic is likely to be the most disastrous public health emergency of our lifetime, with already hundreds of thousands of deaths worldwide. One of the hallmarks of Covid-19 is the wide range of its clinical presentation and ability to affect seemingly healthy patients. Host genetics is therefore a likely determinant of Covid-19 susceptibility and severity. This is supported by recent heritability studies1. Therefore, by studying genetic determinants of Covid-19, we may be able to better understand the biology of Covid-19 severity and susceptibility and subsequently prioritize drug development and clinical trials.

## Scope of the analyses

In this study, we will use WGS or WES data from participants of collaborating cohorts to find genetic determinants of Covid-19 susceptibility, severity, and other associated phenotypes.

A major challenge of WGS and WES data analysis across multiple cohorts is the uniformization of analytic methods to ensure comparable results for the final meta-analysis. A github page with code was made therefore available to guide local cohorts with the analysis, based on the BQC-19 local pipelines, but will still require local adjustments. These are presented as suggestions and may be modified or deviated from based on a local needs assessment. It can be accessed here: <https://github.com/DrGBL/WES.WGS>. We also refer to sequencing quality control pipeline here (sections 1 and 2):

<https://docs.google.com/document/d/1X_qjplH8T4BJXSeMQ_sBfQUTiu_kAisicOqGb6B8hcM/edit>).

# Study Objectives and Endpoints

## Study Objectives

We will perform targeted analysis of WGS and WES data from participants with Covid-19 and clinically similar controls to find genetic variants associated with Covid-19 related phenotypes. Given intrinsic problems with sample size and multiple comparisons affecting statistical power, we will prioritize genes and phenotypes of interest based on previous literature and GWAS findings to date.

## Endpoints, Outcomes, and Measurements

The primary endpoints will follow outcome phenotype definition from the Covid-HGI2. That is, we will study the following three outcomes:

1. Covid-19 susceptibility (C2 in Covid-19 HGI analysis):
   * **Cases**: determined by NAAT (e.g. PCR), serology, medical record review, or self-reported by the patient.
   * **Controls**: every other participants in each cohort that is not a case.
2. Covid-19 hospitalization (B2 in Covid-19 HGI analysis):
   * **Cases**: hospitalized with laboratory confirmed Covid-19. We suggest that a pre-specified timeframe be used to define Covid-19 hospitalization (e.g. from 14 days prior to, and 30 days following the positive test). This can be defined by each site, depending on study design.
   * **Controls**: every other participants in each cohort that is not a case.
3. Severe Covid-19 (A2 in Covid-19 HGI):
   * **Cases**: laboratory confirmed Covid-19 with one or more of the following outcomes:
     1. Death
     2. ECMO requirement
     3. Mechanical ventilation (i.e. intubation) requirement
     4. Non-invasive ventilation requirement (i.e. **new** requirement for BiPAP or CPAP)
     5. High-flow oxygen therapy requirement (e.g. Optiflow)

Again, we suggest that a pre-specified timeframe be used by each study site, depending on study design.

* + **Controls**: every other participants in each cohort that is not a case.

Finally, as secondary endpoints, depending on data availability, we will choose continuous outcomes measured in all participants, and known to be prognostic markers for Covid-19 severe disease or mortality3. This should mitigate problems with lower sample sizes and statistical power. These analyses will only be performed in Covid-19 positive participants. We again suggest that they be measured within a pre-specified timeframe of Covid-19 diagnosis. The outcomes are:

* Maximum circulating neutrophils
* Maximum circulating monocytes
* Minimum circulating lymphocytes
* Minimum circulating CD4 lymphocytes
* Maximum CRP
* Maximum LDH
* Maximum d-dimer
* Maximum PT
* Maximum troponin
* Maximum ferritin
* Maximum SOFA score4

## Genetic loci

To reduce the number of statistical tests performed and preserve as much statistical power as possible, we will restrict our genetic association analysis to pre-specified genome loci known to be involved in the pathophysiology of Covid-19 per the Covid-19 HGI GWAS results. That is, with the exclusion of the human leukocyte antigen region, we will select every protein coding gene within 250 kilobases from the lead SNP of each genome-wide significant loci (i.e. an interval of 500kb in size). The selected genes from the Covid-19 HGI GWAS release 3 are listed in **Appendix 1**.

# Study Methods

## Participants inclusion criteria

Individual participant criteria will depend on the enrolling cohort. For this study, there will be no cohort exclusion criteria based on study design, as long as phenotype outcomes can be measured reliably. Given variations in study design, included cohorts may not be able to provide data about each of the outcomes listed in **Section 4.2**. They may still be included in the study if they are able to perform at least one of the analyses. The only requirement will be informed consent (preferably written) from all study participants (or their surrogate decision makers) to be enrolled in their respective cohort, and to have WGS or WES performed.

## Whole genome or Exome sequencing

There will be no restriction on sequencing technology used as long as it has been properly validated, as per local recommendations. Cohort using either WGS or WES may be included in this study.

## Genetic variant calling and quality control

The GRCh38 reference genome will be used for variant calling and for all downstream analyses. Variants will be joint called locally over all study participants genetic sequencing data. There will be no restriction about genome alignment, quality control, and variant calling pipelines, as long as they follow similar steps and recommendations as GATK Best Practices. If needed, each study site is encouraged to follow the steps outline in our WGS/WES analysis pipeline document (sections 1 and 2 of the following: <https://docs.google.com/document/d/1X_qjplH8T4BJXSeMQ_sBfQUTiu_kAisicOqGb6B8hcM/edit>).

## Genetic variant annotation

From the variant-calling results above, we will perform variant annotation using the following tools, all available using the VEP tool5 (plugins: dbNSFP v4.0 for the first 5 algorithms below, and LOFTEE):

* SIFT6
* PolyPhen27 with HDIV database
* PolyPhen2 with HVAR database
* LRT8
* MutationTaster9
* LoF variants, including stop-gained, essential splice, and frameshift variants will be further annotated using LOFTEE10. LOFTEE implements a set of filters to select variants at high confidence for loss of function. We refer every study site to our WGS/WES analysis pipeline document for more details on LOFTEE.

## Genetic association studies

All analyses will be performed separately for each ancestry, as per each contributing cohort’s population, then meta-analyzed (see below for meta-analysis plan).

The primary analysis will be aligned with previous work done on the UK Biobank whole-exome sequencing from Regeneron’s genetics group11. That is, we will define a burden test by assigning each candidate gene (see **Appendix 1**) per individual a score of 0, 1, or 2 in the following way:

* An individual with no deleterious variant at the gene will receive a score of 0 for that gene.
* An individual with any amount of deleterious variants at the gene, as long as they are all heterozygous, will receive a score of 1 for that gene.
* An individual with any amount of homozygous deleterious variants at the gene will receive a score of 2 for that gene.

This score is similar to an additive effect model for single nucleotide polymorphisms in traditional genome-wide association studies. This is akin to an *in-silico* testing of gene knockdown experiments, where one deleterious variant is assumed to be enough to lead to a loss-of-function, and compounding such variants is hypothesized not to alter function more. To avoid confusion with other types of burden tests, we will refer to this test as the knockdown burden test below.

We will use predicted-deleterious variants according to LOFTEE (a VEP plugin) annotations, and five *in-silico* algorithms to define 4 “masks” to use for our knockdown tests. These five algorithms are the following:

* SIFT
* PolyPhen2 with HDIV database
* PolyPhen2 with HVAR database
* LRT
* MutationTaster

These 5 algorithms are available through the dbSNP4.0a VEP plugin.

All association tests will include the following covariates:

* Age
* Sex
* Age \* Sex
* Age squared
* Age squared \* sex
* 10 PCs obtained from common variants (e.g. from whole-genome genotyping, as in GWASs), using a MAF lower bound of at least 1%.
* 20 PCs obtained from rare variants12, using LD pruned variants with minor allele frequency less than 1% (allele frequency based on each individual cohort’s sample). Lower threshold on MAC to be left at the discretion of each participating cohort.

The association studies will therefore be performed locally in two ways:

1. Knockdown burden tests:
   * The following four masks will be used (in separate analyses), on each gene’s *canonical transcript*:
     1. M1: using pLOF variants (high confidence [HC] LOF per LOFTEE)
     2. M2: using pLOF variants OR any missense variants (e.g. annotated “missense\_variant” by VEP)
     3. M3: using pLOF variants OR missense variants classified as deleterious by all five *in-silico* algorithms above.
     4. M4: using pLOF variants OR missense variants classified as deleterious by at least one of the five *in-silico* algorithms above.
   * To ensure that all cohorts use a common definition of what constitutes a MAF below 1%, we will only allow the use of the following variants (assuming they are selected using the masks above):
     1. Variants present in gnomAD,1000G, or ESP with MAF<1% (can be obtained with the --max-af option in VEP)
     2. Variants not present in gnomAD,1000G, or ESP but with a MAF<1% in a specific cohort.

In further analysis, and depending on sample size, the MAF threshold will be lowered to restrict to ultra-rare variants.

* + P-value threshold for knockdown burden tests will be obtained using Bonferroni adjustments by the number of genes included, with a significant p-value of 0.05, divided by the number of genes tested.

1. Individual deleterious variants association tests.
   * If we find significant associations using gene-based tests, we will do single association studies using all variants from those genes.
   * Power curves based on the hospitalized phenotypes, with a significance threshold of 5x10-8, and using the November 17 number of cases reported in the consortium were performed using the Genetics Association Study Power Calculator13. They are shown in **Appendix 2**. These suggests that we should be well powered for relative risks > 2:
   * P-value threshold for these tests will be obtained using Bonferroni adjustments by the number of variants tested (note that this is a likely conservative estimate, given linkage disequilibrium).

All analyses will be performed using the Regenie software14. Regenie uses a two-step approach to performing genome-wise association studies:

* + Step 1: uses ridge regression to produce a genome-wide prediction for the outcomes. Common variants (i.e. like a GWAS) should be used for this step. Step 1 only needs to be done 1, and multiple phenotypes can be combined at once.
  + Step 2: using results from Step 1, we will analyze each gene and mask combinations as though they were a single nucleotide polymorphism, using the knock-down burden test described above. This step uses Firth penalized likelihood to adjust for rare events.

## Meta-analysis

Summary statistics from the knockdown burden will be used in an inverse-variance weighted random effect model, to account for the expected ancestry heterogeneity. This will be done separately for every mask in the knockout tests. The same procedure will be used if single variant association tests are done. The following results will be needed for summary statistics, some of which are already available from the standard Regenie result output:

* + mask used
  + gene (recommend ensemble gene ID for ease)
  + “Effect” allele used for the gene-SNP model in Regenie
  + “Other” allele
  + beta from knockdown burden test
  + standard error
  + p-value
  + sample size
  + ancestry
  + MAF upper bound

# Data sharing

Meta-analysis results will be made available freely to researchers without restrictions. For access to individual cohort or individual patient data, researchers will need to apply for data access as per each cohort’s individual data sharing rules.

# References

1. Williams FMK, Freydin M, Mangino M, et al. Self-reported symptoms of covid-19 including symptoms most predictive of SARS-CoV-2 infection, are heritable. *medRxiv*. Published online January 1, 2020:2020.04.22.20072124. doi:10.1101/2020.04.22.20072124

2. Covid-19 Host Genetic Initiative Phenotype definitions for analyses. Published 2020. Accessed September 9, 2020. https://docs.google.com/document/d/1okamrqYmJfa35ClLvCt\_vEe4PkvrTwggHq7T3jbeyCI/edit

3. Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A. The role of biomarkers in diagnosis of COVID-19 - A systematic review. *Life Sci*. 2020;254:117788. doi:10.1016/j.lfs.2020.117788

4. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score—development, utility and challenges of accurate assessment in clinical trials. *Crit Care*. 2019;23(1):374. doi:10.1186/s13054-019-2663-7

5. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4

6. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11(1):1-9. doi:10.1038/nprot.2015.123

7. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20. doi:10.1002/0471142905.hg0720s76

8. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009;19(9):1553-1561. doi:10.1101/gr.092619.109

9. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575-576. doi:10.1038/nmeth0810-575

10. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

11. Kosmicki JA, Horowitz JE, Banerjee N, et al. Genetic association analysis of SARS-CoV-2 infection in 455,838 UK Biobank participants. *medRxiv*. Published online January 1, 2020:2020.10.28.20221804. doi:10.1101/2020.10.28.20221804

12. Mathieson I, McVean G. Differential confounding of rare and common variants in spatially structured populations. *Nat Genet*. 2012;44(3):243-246. doi:10.1038/ng.1074

13. Johnson JL, Abecasis GR. GAS Power Calculator: web-based power calculator for genetic association studies. *bioRxiv*. Published online January 1, 2017:164343. doi:10.1101/164343

14. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv*. Published online January 1, 2020:2020.06.19.162354. doi:10.1101/2020.06.19.162354

# Appendix 1

Genes selected for the three analyses. Gene positions are given using the GRCh38 reference genome. Genes listed by chromosome and start positions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Covid-19 Susceptibility Analysis** | | | | | |
| **HGCN Symbol** | **Ensembl Gene ID** | **Description** | **Chr** | **Start** | **End** |
| LIMD1 | ENSG00000144791 | LIM domains containing 1 | 3 | 45555394 | 45686341 |
| SACM1L | ENSG00000211456 | SAC1 like phosphatidylinositide phosphatase | 3 | 45689056 | 45745409 |
| SLC6A20 | ENSG00000163817 | solute carrier family 6 member 20 | 3 | 45755449 | 45796536 |
| LZTFL1 | ENSG00000163818 | leucine zipper transcription factor like 1 | 3 | 45823316 | 45916042 |
| CCR9 | ENSG00000173585 | C-C motif chemokine receptor 9 | 3 | 45886504 | 45903175 |
| FYCO1 | ENSG00000163820 | FYVE and coiled-coil domain autophagy adaptor 1 | 3 | 45917899 | 45995824 |
| CXCR6 | ENSG00000172215 | C-X-C motif chemokine receptor 6 | 3 | 45940933 | 45948351 |
| XCR1 | ENSG00000173578 | X-C motif chemokine receptor 1 | 3 | 46017024 | 46027742 |
| TLE1 | ENSG00000196781 | TLE family member 1, transcriptional corepressor | 9 | 81583683 | 81689547 |
| GTF3C5 | ENSG00000148308 | general transcription factor IIIC subunit 5 | 9 | 1.33E+08 | 1.33E+08 |
| CEL | ENSG00000170835 | carboxyl ester lipase | 9 | 1.33E+08 | 1.33E+08 |
| RALGDS | ENSG00000160271 | ral guanine nucleotide dissociation stimulator | 9 | 1.33E+08 | 1.33E+08 |
| GBGT1 | ENSG00000148288 | globoside alpha-1,3-N-acetylgalactosaminyltransferase 1 (FORS blood group) | 9 | 1.33E+08 | 1.33E+08 |
| OBP2B | ENSG00000171102 | odorant binding protein 2B | 9 | 1.33E+08 | 1.33E+08 |
| ABO | ENSG00000175164 | ABO, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase | 9 | 1.33E+08 | 1.33E+08 |
| SURF6 | ENSG00000148296 | surfeit 6 | 9 | 1.33E+08 | 1.33E+08 |
| MED22 | ENSG00000148297 | mediator complex subunit 22 | 9 | 1.33E+08 | 1.33E+08 |
| RPL7A | ENSG00000148303 | ribosomal protein L7a | 9 | 1.33E+08 | 1.33E+08 |
| SURF1 | ENSG00000148290 | SURF1 cytochrome c oxidase assembly factor | 9 | 1.33E+08 | 1.33E+08 |
| SURF2 | ENSG00000148291 | surfeit 2 | 9 | 1.33E+08 | 1.33E+08 |
| SURF4 | ENSG00000148248 | surfeit 4 | 9 | 1.33E+08 | 1.33E+08 |
| STKLD1 | ENSG00000198870 | serine/threonine kinase like domain containing 1 | 9 | 1.33E+08 | 1.33E+08 |
| REXO4 | ENSG00000148300 | REX4 homolog, 3'-5' exonuclease | 9 | 1.33E+08 | 1.33E+08 |
| ADAMTS13 | ENSG00000160323 | ADAM metallopeptidase with thrombospondin type 1 motif 13 | 9 | 1.33E+08 | 1.33E+08 |
| CACFD1 | ENSG00000160325 | calcium channel flower domain containing 1 | 9 | 1.33E+08 | 1.33E+08 |
| SLC2A6 | ENSG00000160326 | solute carrier family 2 member 6 | 9 | 1.33E+08 | 1.33E+08 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Covid-19 Hospitalisation Analysis** | | | | | |
| **HGCN Symbol** | **Ensembl Gene ID** | **Description** | **Chr** | **Start** | **End** |
| LIMD1 | ENSG00000144791 | LIM domains containing 1 | 3 | 45555394 | 45686341 |
| SACM1L | ENSG00000211456 | SAC1 like phosphatidylinositide phosphatase | 3 | 45689056 | 45745409 |
| SLC6A20 | ENSG00000163817 | solute carrier family 6 member 20 | 3 | 45755449 | 45796536 |
| LZTFL1 | ENSG00000163818 | leucine zipper transcription factor like 1 | 3 | 45823316 | 45916042 |
| CCR9 | ENSG00000173585 | C-C motif chemokine receptor 9 | 3 | 45886504 | 45903175 |
| FYCO1 | ENSG00000163820 | FYVE and coiled-coil domain autophagy adaptor 1 | 3 | 45917899 | 45995824 |
| CXCR6 | ENSG00000172215 | C-X-C motif chemokine receptor 6 | 3 | 45940933 | 45948351 |
| XCR1 | ENSG00000173578 | X-C motif chemokine receptor 1 | 3 | 46017024 | 46027742 |
| TREM1 | ENSG00000124731 | triggering receptor expressed on myeloid cells 1 | 6 | 41267926 | 41286682 |
| NCR2 | ENSG00000096264 | natural cytotoxicity triggering receptor 2 | 6 | 41335608 | 41350889 |
| FOXP4 | ENSG00000137166 | forkhead box P4 | 6 | 41546426 | 41602384 |
| MDFI | ENSG00000112559 | MyoD family inhibitor | 6 | 41636882 | 41654246 |
| TFEB | ENSG00000112561 | transcription factor EB | 6 | 41683978 | 41736259 |
| PGC | ENSG00000096088 | progastricsin | 6 | 41736711 | 41754109 |
| FRS3 | ENSG00000137218 | fibroblast growth factor receptor substrate 3 | 6 | 41770176 | 41786542 |
| RPH3A | ENSG00000089169 | rabphilin 3A | 12 | 1.13E+08 | 1.13E+08 |
| OAS1 | ENSG00000089127 | 2'-5'-oligoadenylate synthetase 1 | 12 | 1.13E+08 | 1.13E+08 |
| OAS3 | ENSG00000111331 | 2'-5'-oligoadenylate synthetase 3 | 12 | 1.13E+08 | 1.13E+08 |
| OAS2 | ENSG00000111335 | 2'-5'-oligoadenylate synthetase 2 | 12 | 1.13E+08 | 1.13E+08 |
| DTX1 | ENSG00000135144 | deltex E3 ubiquitin ligase 1 | 12 | 1.13E+08 | 1.13E+08 |
| RASAL1 | ENSG00000111344 | RAS protein activator like 1 | 12 | 1.13E+08 | 1.13E+08 |
| CFAP73 | ENSG00000186710 | cilia and flagella associated protein 73 | 12 | 1.13E+08 | 1.13E+08 |
| DDX54 | ENSG00000123064 | DEAD-box helicase 54 | 12 | 1.13E+08 | 1.13E+08 |
| HDGFL2 | ENSG00000167674 | HDGF like 2 | 19 | 4472297 | 4502208 |
| PLIN4 | ENSG00000167676 | perilipin 4 | 19 | 4502180 | 4518465 |
| PLIN5 | ENSG00000214456 | perilipin 5 | 19 | 4522531 | 4535224 |
| LRG1 | ENSG00000171236 | leucine rich alpha-2-glycoprotein 1 | 19 | 4536402 | 4540036 |
| SEMA6B | ENSG00000167680 | semaphorin 6B | 19 | 4542593 | 4559684 |
| TNFAIP8L1 | ENSG00000185361 | TNF alpha induced protein 8 like 1 | 19 | 4639516 | 4655568 |
| MYDGF | ENSG00000074842 | myeloid derived growth factor | 19 | 4641374 | 4670362 |
| DPP9 | ENSG00000142002 | dipeptidyl peptidase 9 | 19 | 4675224 | 4724673 |
| FEM1A | ENSG00000141965 | fem-1 homolog A | 19 | 4791734 | 4801273 |
| TICAM1 | ENSG00000127666 | toll like receptor adaptor molecule 1 | 19 | 4815932 | 4831712 |
| PLIN3 | ENSG00000105355 | perilipin 3 | 19 | 4838341 | 4867694 |
| ARRDC5 | ENSG00000205784 | arrestin domain containing 5 | 19 | 4890437 | 4902896 |
| UHRF1 | ENSG00000276043 | ubiquitin like with PHD and ring finger domains 1 | 19 | 4903080 | 4962154 |
| KDM4B | ENSG00000127663 | lysine demethylase 4B | 19 | 4969113 | 5153598 |
| SHFL | ENSG00000130813 | shiftless antiviral inhibitor of ribosomal frameshifting | 19 | 10086122 | 10093252 |
| ANGPTL6 | ENSG00000130812 | angiopoietin like 6 | 19 | 10092338 | 10102796 |
| PPAN-P2RY11 | ENSG00000243207 | PPAN-P2RY11 readthrough | 19 | 10106223 | 10114780 |
| PPAN | ENSG00000130810 | peter pan homolog | 19 | 10106362 | 10112012 |
| P2RY11 | ENSG00000244165 | purinergic receptor P2Y11 | 19 | 10111693 | 10115372 |
| EIF3G | ENSG00000130811 | eukaryotic translation initiation factor 3 subunit G | 19 | 10115014 | 10119918 |
| DNMT1 | ENSG00000130816 | DNA methyltransferase 1 | 19 | 10133345 | 10231286 |
| S1PR2 | ENSG00000267534 | sphingosine-1-phosphate receptor 2 | 19 | 10221433 | 10231331 |
| MRPL4 | ENSG00000105364 | mitochondrial ribosomal protein L4 | 19 | 10251901 | 10260055 |
| ICAM1 | ENSG00000090339 | intercellular adhesion molecule 1 | 19 | 10271093 | 10286615 |
| ICAM4 | ENSG00000105371 | intercellular adhesion molecule 4 (Landsteiner-Wiener blood group) | 19 | 10286955 | 10288522 |
| ICAM5 | ENSG00000105376 | intercellular adhesion molecule 5 | 19 | 10289952 | 10296778 |
| ZGLP1 | ENSG00000220201 | zinc finger GATA like protein 1 | 19 | 10304803 | 10309880 |
| FDX2 | ENSG00000267673 | ferredoxin 2 | 19 | 10310045 | 10316015 |
| RAVER1 | ENSG00000161847 | ribonucleoprotein, PTB binding 1 | 19 | 10316212 | 10333638 |
| ICAM3 | ENSG00000076662 | intercellular adhesion molecule 3 | 19 | 10333776 | 10339661 |
| TYK2 | ENSG00000105397 | tyrosine kinase 2 | 19 | 10350529 | 10380572 |
| CDC37 | ENSG00000105401 | cell division cycle 37, HSP90 cochaperone | 19 | 10391133 | 10420121 |
| PDE4A | ENSG00000065989 | phosphodiesterase 4A | 19 | 10416773 | 10469630 |
| KEAP1 | ENSG00000079999 | kelch like ECH associated protein 1 | 19 | 10486125 | 10503558 |
| S1PR5 | ENSG00000180739 | sphingosine-1-phosphate receptor 5 | 19 | 10512742 | 10517931 |
| ATG4D | ENSG00000130734 | autophagy related 4D cysteine peptidase | 19 | 10543895 | 10553418 |
| KRI1 | ENSG00000129347 | KRI1 homolog | 19 | 10553085 | 10566031 |
| CDKN2D | ENSG00000129355 | cyclin dependent kinase inhibitor 2D | 19 | 10566460 | 10569059 |
| OLIG2 | ENSG00000205927 | oligodendrocyte transcription factor 2 | 21 | 33025935 | 33029196 |
| OLIG1 | ENSG00000184221 | oligodendrocyte transcription factor 1 | 21 | 33070141 | 33072413 |
| IFNAR2 | ENSG00000159110 | interferon alpha and beta receptor subunit 2 | 21 | 33229901 | 33265675 |
| IL10RB | ENSG00000243646 | interleukin 10 receptor subunit beta | 21 | 33266367 | 33310187 |
| IFNAR1 | ENSG00000142166 | interferon alpha and beta receptor subunit 1 | 21 | 33324429 | 33359864 |
| IFNGR2 | ENSG00000159128 | interferon gamma receptor 2 | 21 | 33403413 | 33479348 |
| TMEM50B | ENSG00000142188 | transmembrane protein 50B | 21 | 33432485 | 33479974 |
| DNAJC28 | ENSG00000177692 | DnaJ heat shock protein family (Hsp40) member C28 | 21 | 33485530 | 33491716 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Covid-19 Severe Disease Analysis** | | | | | |
| **HGCN Symbol** | **Ensembl Gene ID** | **Description** | **Chr** | **Start** | **End** |
| LIMD1 | ENSG00000144791 | LIM domains containing 1 | 3 | 45555394 | 45686341 |
| SACM1L | ENSG00000211456 | SAC1 like phosphatidylinositide phosphatase | 3 | 45689056 | 45745409 |
| SLC6A20 | ENSG00000163817 | solute carrier family 6 member 20 | 3 | 45755449 | 45796536 |
| LZTFL1 | ENSG00000163818 | leucine zipper transcription factor like 1 | 3 | 45823316 | 45916042 |
| CCR9 | ENSG00000173585 | C-C motif chemokine receptor 9 | 3 | 45886504 | 45903175 |
| FYCO1 | ENSG00000163820 | FYVE and coiled-coil domain autophagy adaptor 1 | 3 | 45917899 | 45995824 |
| CXCR6 | ENSG00000172215 | C-X-C motif chemokine receptor 6 | 3 | 45940933 | 45948351 |
| XCR1 | ENSG00000173578 | X-C motif chemokine receptor 1 | 3 | 46017024 | 46027742 |
| VSTM2A | ENSG00000170419 | V-set and transmembrane domain containing 2A | 7 | 54542325 | 54571080 |
| SEC61G | ENSG00000132432 | SEC61 translocon subunit gamma | 7 | 54752250 | 54759974 |
| IGF1 | ENSG00000017427 | insulin like growth factor 1 | 12 | 1.02E+08 | 1.02E+08 |
| PAH | ENSG00000171759 | phenylalanine hydroxylase | 12 | 1.03E+08 | 1.03E+08 |
| RPH3A | ENSG00000089169 | rabphilin 3A | 12 | 1.13E+08 | 1.13E+08 |
| OAS1 | ENSG00000089127 | 2'-5'-oligoadenylate synthetase 1 | 12 | 1.13E+08 | 1.13E+08 |
| OAS3 | ENSG00000111331 | 2'-5'-oligoadenylate synthetase 3 | 12 | 1.13E+08 | 1.13E+08 |
| OAS2 | ENSG00000111335 | 2'-5'-oligoadenylate synthetase 2 | 12 | 1.13E+08 | 1.13E+08 |
| DTX1 | ENSG00000135144 | deltex E3 ubiquitin ligase 1 | 12 | 1.13E+08 | 1.13E+08 |
| RASAL1 | ENSG00000111344 | RAS protein activator like 1 | 12 | 1.13E+08 | 1.13E+08 |
| CFAP73 | ENSG00000186710 | cilia and flagella associated protein 73 | 12 | 1.13E+08 | 1.13E+08 |
| DDX54 | ENSG00000123064 | DEAD-box helicase 54 | 12 | 1.13E+08 | 1.13E+08 |
| RITA1 | ENSG00000139405 | RBPJ interacting and tubulin associated 1 | 12 | 1.13E+08 | 1.13E+08 |
| HDGFL2 | ENSG00000167674 | HDGF like 2 | 19 | 4472297 | 4502208 |
| PLIN4 | ENSG00000167676 | perilipin 4 | 19 | 4502180 | 4518465 |
| PLIN5 | ENSG00000214456 | perilipin 5 | 19 | 4522531 | 4535224 |
| LRG1 | ENSG00000171236 | leucine rich alpha-2-glycoprotein 1 | 19 | 4536402 | 4540036 |
| SEMA6B | ENSG00000167680 | semaphorin 6B | 19 | 4542593 | 4559684 |
| TNFAIP8L1 | ENSG00000185361 | TNF alpha induced protein 8 like 1 | 19 | 4639516 | 4655568 |
| MYDGF | ENSG00000074842 | myeloid derived growth factor | 19 | 4641374 | 4670362 |
| DPP9 | ENSG00000142002 | dipeptidyl peptidase 9 | 19 | 4675224 | 4724673 |
| FEM1A | ENSG00000141965 | fem-1 homolog A | 19 | 4791734 | 4801273 |
| TICAM1 | ENSG00000127666 | toll like receptor adaptor molecule 1 | 19 | 4815932 | 4831712 |
| PLIN3 | ENSG00000105355 | perilipin 3 | 19 | 4838341 | 4867694 |
| ARRDC5 | ENSG00000205784 | arrestin domain containing 5 | 19 | 4890437 | 4902896 |
| UHRF1 | ENSG00000276043 | ubiquitin like with PHD and ring finger domains 1 | 19 | 4903080 | 4962154 |
| KDM4B | ENSG00000127663 | lysine demethylase 4B | 19 | 4969113 | 5153598 |
| OLIG2 | ENSG00000205927 | oligodendrocyte transcription factor 2 | 21 | 33025935 | 33029196 |
| OLIG1 | ENSG00000184221 | oligodendrocyte transcription factor 1 | 21 | 33070141 | 33072413 |
| IFNAR2 | ENSG00000159110 | interferon alpha and beta receptor subunit 2 | 21 | 33229901 | 33265675 |
| IL10RB | ENSG00000243646 | interleukin 10 receptor subunit beta | 21 | 33266367 | 33310187 |
| IFNAR1 | ENSG00000142166 | interferon alpha and beta receptor subunit 1 | 21 | 33324429 | 33359864 |
| IFNGR2 | ENSG00000159128 | interferon gamma receptor 2 | 21 | 33403413 | 33479348 |
| TMEM50B | ENSG00000142188 | transmembrane protein 50B | 21 | 33432485 | 33479974 |
| DNAJC28 | ENSG00000177692 | DnaJ heat shock protein family (Hsp40) member C28 | 21 | 33485530 | 33491716 |

# Appendix 2

Power calculations using the Genetics Association Study Power Calculation Tool: <http://csg.sph.umich.edu/abecasis/gas_power_calculator/>

