**Supplementary Tables**

**Table S1.** Symptoms with respective cut-off values used in the Generation Scotland, Helix, Lifelines, and NTR cohorts. Logistic regression was performed for each symptom separately in Lifelines on positive (n=56) vs negative (n=586) tested individuals to define symptom cut-offs. The odds ratios and p-values of used cut-offs in Lifelines are displayed here.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Symptom** | **Generation Scotland** | **Helix** | **Lifelines** | **NTR** |
| Timeframe over which symptoms were questioned | No specified timeframe, could be either past or current symptoms | Any time since Jan 1 2020, with approximate duration of symptoms (wave) | Last 7 days (weekly waves)For tested subjects we used the maximum weekly symptom score from the week of the test and the 3 weeks prior to the test, for not tested subjects we used the maximum weekly symptom score reported in the first 7 weeks of the study. | Any time since February 20th 2020 (two waves) |
|  | Scale | Scale | Scale | Cut-off to define positive symptom  | OR (p-value) | Scale | Cut-off to define positive symptom  |
| Aches and pains | Binary | NA | NA | NA | NA | NA | NA |
| Back pain | NA | Binary | 5-point | ≥ 3 | 2.40 (3.3\*10-3) | 5-point\*\*\* | ≥ 3 |
| Chest pain | NA | Binary | 5-point | ≥ 2 | 1.91 (2.1\*10-2) | 5-point\*\*\* | ≥ 2 |
| Coughing - any | NA | NA | 5-pointMaximum value of cough with and without sputum | ≥ 3 (Menni)≥ 2 (Lifelines) | 6.53 (3.6\*10-4) | 5-point | ≥ 4 (Menni)≥ 2 (Lifelines) |
| Coughing - with sputum | NA | Binary | 5-point | ≥ 2 | 1.37 (2.6\*10-1) | NA | NA |
| Coughing - without sputum (i.e. dry cough) | Binary | Binary | 5-point | ≥ 2 | 5.14 (7.3\*10-5) | NA | NA |
| Decrease in appetite | NA | Binary | NA | NA | NA | NA | NA |
| Developed pneumonia | Binary | NA | NA | NA | NA | NA | NA |
| Diarrhea | Binary | Binary | 5-point | ≥ 2 | 2.88 (1.8\*10-4) | NA | NA |
| Diarrhea or stomach ache | NA | NA | 5-point\* | ≥ 2  | 2.05 (1.1\*10-2) | 5-point | ≥ 2 |
| Difficulty breathing | NA | Binary | 5-point | ≥ 2 | 2.66 (7.2\*10-4) | 5-point | ≥ 2 |
| Dizziness | NA | Binary | 5-point | ≥ 3  | 3.42 (3.1\*10-4) | 5-point\*\*\* | ≥ 3 |
| Feeling fit | NA | Binary | left-anchored 7-point | ≥ 6 | 3.93 (8.8\*10-6) | NA | NA |
| Feeling heavy in arms/legs | NA | Binary | 5-point | ≥ 3  | 3.98 (1.5\*10-6) | NA | NA |
| Feeling physically exhausted | NA | Binary | left-anchored 7-point | ≤ 2 | 3.54 (9.2\*10-6) | NA | NA |
| Feeling tired | Binary (fatigue/ tiredness) | NA | left-anchored 7-point | ≤ 2 | 4.68 (5.5\*10-7) | 5-point (felt tired or limp) | > 4 (Menni)≥ 2 (Lifelines) |
| Feeling tired quickly | NA | Binary | left-anchored 7-point | ≤ 2 | 4.94 (2.3\*10-7) | NA | NA |
| Feeling limp | NA | Binary | 5-point | ≥ 3  | 7.26 (1.2\*10-10) | NA | NA |
| Fever | Binary (fever/high temperature) | Binary | 5-point | ≥ 2 | 13.60 (1.8\*10-16) | 5-point | ≥ 2 |
| Headache | Binary | Binary | 5-point | ≥ 3  | 9.32 (9.4\*10-13) | 5-point\*\*\* | ≥ 3 |
| Lack of appetite | Binary | NA | NA | NA | NA | NA | NA |
| Loss of appetite | NA | NA | NA | NA | NA | 5-point | ≥4 (Menni) |
| Loss of smell/taste | Binary | Binary | 5-point | ≥ 3 | 17.90 (6.5\*10-19) | 5-point | ≥ 2 (Menni)≥ 3 (Lifelines) |
| Lump in throat | NA | Binary | 5-point | ≥ 2 | 2.87 (3.0\*10-4) | NA | NA |
| Muscle soreness | NA | Binary | 5-point | ≥ 3  | 3.35 (2.1\*10-5) | 5-point\*\*\* | ≥ 3  |
| Nausea | Binary (nausea/ feeling sick) | Binary | 5-point | ≥ 2 | 2.12 (9.1\*10-3) | 5-point\*\*\* | ≥ 2 |
| Numbness/ tinglings | NA | Binary | 5-point | ≥ 3  | 1.95 (5.6\*10-2) | NA | NA |
| Pain when breathing | NA | Binary (pain with deep breaths) | 5-point | ≥ 2 | 3.30 (3.4\*10-5) | 5-point\*\*\* | ≥ 2 |
| Phlegm | NA | Binary | NA | NA | NA | NA | NA |
| Red/painful/itchy eyes | NA | Binary | 5-point | ≥ 2 | 1.40 (2.3\*10-1) | NA | NA |
| Runny nose | Binary | Binary | 5-point | ≥ 3 | 2.65 (7.8\*10-4) | 5-point (common/nasal cold) | ≥ 3 |
| Sensitive skin | NA | Binary | NA | NA | NA | NA | NA |
| Severe fatigue, such as inability to get out of bed | NA | Binary | NA | NA | NA | NA | NA |
| Shortness of breath | Binary | NA | 5-point | ≥ 3  | 2.58 (1.6\*10-3) | NA | NA |
| Sneezing | NA | NA | 5-point\*\* | ≥ 3 | 1.66 (1.3\*10-1) | NA | NA |
| Sore eyes | Binary | NA | NA | NA | NA | NA | NA |
| Sore throat | Binary | Binary | 5-point | ≥ 2 | 2.09 (1.6\*10-2) | 5-point (sore throat/laryngitis) | ≥ 2 |
| Stomach ache | Binary | NA | 5-point | ≥ 2  | 2.17 (6.6\*10-3) | NA | NA |
| Temperature fluctuations | NA | Binary | 5-point | ≥ 2  | 12.66 (1.4\*10-6) | NA | NA |
| Tired - any | NA | Binary | left-anchored 7-pointMinimum value of feeling tired and feeling tired quickly (for the Menni model: also feeling physically exhausted) | ≤ 2 | 4.85 (8.2\*10-7) | NA | NA |

\* Questioned from week 3 of the Lifelines cohort

\*\* Questioned from week 2 of the Lifelines cohort

\*\*\* Questioned in the second wave only

**Table S2**. The different cut-offs and symptoms that were used by the Generation Scotland, Helix, Lifelines and NTR cohorts when applying the Menni COVID-19 prediction model to the datasets prior to running the GWAS.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Symptom in Menni model** | **Generation Scotland** | **Helix** | **Lifelines** | **NTR** |
| **Loss of smell and taste** | Binary | Binary | 5-point, ≥ 3 | 5-point, ≥ 2 |
| **Severe or significant persistent cough** | Dry cough, binary | Dry cough lasting at least 20 days, binary | Maximum value of cough with and cough without sputum, 5-point, ≥ 3 | Any cough, 5-point, ≥ 4 |
| **Severe fatigue** | Fatigue/ tiredness, binary | Severe fatigue, such as inability to get out of bed, Binary | Minimum of feeling tired, feeling tired quickly, and feeling physically exhausted, left-anchored 7-point, ≤ 2 | Feeling tired or limp, 5-point, > 4 |
| **Skipped meals** | Lack of appetite, binary | Decrease in appetite, Binary | NA | Loss of appetite, 5-point, ≥ 4 |

**Table S3.** Full phenotype description and detailed analysis plan used within the C19HG for genome-wide association analysis

**27 April 2020 - version 2.0**

For V1.0 of this document, used for first two rounds of GWAS meta-analyses, please view [her](https://docs.google.com/document/d/1eMdzhO5xk-MACxjz-kOUJLP6Jort5KuwoOa_u-aZPHs/edit?usp=sharing)​ [e](https://docs.google.com/document/d/1eMdzhO5xk-MACxjz-kOUJLP6Jort5KuwoOa_u-aZPHs/edit?usp=sharing)

**Working group**

**Lea Davis,** expertise: genetic medicine, human genomics, EHR based genomics​

**Andrea Ganna,** expertise: statistical genetics, epidemiology​

**Sulggi Lee,** expertise: infectious disease physician, covid research, enrolling a cohort​

**James Priest,** expertise: paed cardiologist, GWAS, health informatics.​

**Alessandra Renieri,** expertise: prof of medical genetics, WES, consultant physician (rare​ diseases), CAT member of EMA.

**Vijay Sankaran,** expertise: paed heamatol/oncol, human geneticist, blood cell traits​

**David van Heel,** expertise: prof of human genetics (Genes & Health cohort CI), consultant​ physician (adults) Barts Health NHS Trust, deputy Chief Clinical Information Officer, clinical IT on covid patients (e.g. for Nightingale & Barts Health).

**Patrick Deelen**, expertise: postdoc Lifelines (Lude Franke), bioinformatician human genetics​

**Brent Richards**, expertise: prof of medicine, endocrinology, seeing covid patients, genetic​ epidemiology.

**Tomoko Nakanishi:** respiratory doctor, PhD.​

**Les Biesecker**, paed doctor, medical geneticist, PI covid cohort US NIH (recruiting)​

**Eric Kerchberger**:​ pulmonary clinical fellow, MD. Critical illness syndromes.

**RATIONALE:**​​Identification of human genetic factors (should these exist) impacting the COVID19 pandemic holds the potential to rapidly offer important information. Within individual cohorts, we propose establishing a shared set of phenotypes to be combined with genomic data for standard GWAS and further meta-analysis by the analysis group. Analyses will be simple and scalable.



**Analysis A:** ​ ​**VERY**​ ​**SEVERE RESPIRATORY CONFIRMED COVID**

Analysis A1\_v2 ​(name ANA\_A1\_V2) ​-- PRIMARY ANALYSIS

**Cases:**​ ​Hospitalized laboratory confirmed​ SARS-CoV-2 infection (RNA and/or serology based), AND (​death​ OR ​respiratory support​ (intubation, CPAP, BiPAP, CNP (continue external negative pressure), Optiflow/very high flow Positive End Expiratory Pressure Oxygen\* - (see ​**Appendix 1** for suggestive codes) AND hospitalization with COVID19 as primary reason for admission..

\*Simple supplementary oxygen (e.g. 2 litres/minute via nasal cannulae) ​only​ does not qualify for case status.

**Controls:** ​Laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based) AND not hospitalised 21 days after the test.

**Analysis A2\_v2**​​(name ANA\_A2\_V2)

**Cases:**​ As A1\_v2

**Controls:** ​everybody that is not a case, e.g. population

**Analysis B:** ​**HOSPITALIZED LAB CONFIRMED COVID**

Analysis B1\_v2 ​(name ANA\_B1\_V2) ​-- PRIMARY ANALYSIS

**Cases:**​ ​Hospitalized laboratory confirmed​ SARS-CoV-2 infection (RNA and/or serology based), hospitalization ​**due to corona-related symptoms**​.

**Controls:** ​Laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based) AND not hospitalised 21 days after the test.

**Analysis B2\_v2**​​(name ANA\_B2\_V2)

**Cases:**​ As B1\_v2

**Controls:** ​everybody that is not a case, e.g. population

**Analysis B3\_v2** ​(name ANA\_B3\_V2)

**Cases:**​ ​Hospitalized laboratory confirmed​ SARS-CoV-2 infection (RNA and/or serology based), hospitalization due to corona-related symptoms OR ​self-reported hospitalized COVID-19 positive​ (e.g. by questionnaire)

**Controls:** ​(Laboratory confirmed SARS-CoV-2 infection (RNA and/or serology based) AND not hospitalised 21 days after the test) OR ​self-reported COVID-19 positive without hospitalization **Analysis C:** ​**PARTIAL**​-​**SUSCEPTIBILITY**

**Analysis C1\_v2** ​(name ANA\_C1\_V2)

**Cases:**​ individuals with ​laboratory confirmation​ of SARS-CoV-2 infection (RNA and/or serology based) OR ​EHR/ICD coding/ Physician Confirmed​ COVID-19 (See ​**Appendix 1**​ for suggestive codes) OR ​self-reported COVID-19 positive​ (e.g. by questionnaire)

**Controls:** ​(Laboratory tested for SARS-CoV-2 infection (RNA and/or serology based) AND all tests (if multiple tests) ​negative​\*) OR ​self-reported​ tested negative for SARS-CoV-2 infection (e.g. by questionnaire)

\*​We suggest to exclude from the control group individuals with high clinical suspicion of COVID-19 but negative laboratory tests

# Analysis C2\_v2​ ​(name ANA\_C2\_V2)​ -- PRIMARY ANALYSIS

**Cases:** ​As C1\_v2

**Controls:** ​everybody that is not a case, e.g. population

**Analysis D: COVID-19 PREDICTED FROM SELF-REPORTED SYMPTOMS**

**Analysis D1\_v2 (**name ANA\_D1\_V2​ **)**​

**Cases:** Samples with a value from the predictive model\* > -0.44 OR ​ self-reported COVID-19​ positive

\*The model from ​[Menni et al. 2020](https://www.nature.com/articles/s41591-020-0916-2)​, minus age and sex predictors, will be applied to self-report data. The published model is as follows:

Predictionmodel=−1.32+(1.75×lossofsmellandtaste)+(0.31×severeorsignificantpersistentcough)+(0.49×severefatigue)

+(0.39×skippedmeals) where all symptoms are coded as 1 if the person self-reports the symptom and 0 if not.

The AUC for the original version of this model, which included age and sex, has been shown in both the Helix and Lifelines dataset to be 0.79 despite slight differences in the phenotypes (for example, skipped meals were not included in the Lifelines model, and Helix used any dry cough lasting at least 20 days as a predictor instead of self-reported severe or persistant cough). In the Helix dataset, we have observed that removing age and sex from the model, which makes more sense for genetic analyses, results in minimal changes of the AUC (to 0.78). In the Helix dataset, the final model produces PPV of 0.41, sensitivity 0.62, and specificity 0.88.

**Controls:** Samples with the minimum possible value from the predictive model (i.e., value from​ the predictive model = -1.32) AND NOT self-reported COVID-19 positive

Please note that **age, age^2, sex and age\*sex are included as a covariate in all analyses**​  **A,B,C,D above. Therefore these analyses already consider, to some extent, the difference between younger and older (with e.g. more comorbidities) patients.**

# ANALYSIS PHENOTYPES FOR CONSIDERATION IN V3

Mortality in young individuals

Min lymphocyte levels during hospitalization

Max granulocyte and monocyte levels during hospitalization

Max IL-6 levels during hospitalization

Max CRP levels during hospitalization

Min Fibrinogen levels during hospitalization

Severity scale during hospitalization

Loss of smell or taste

Neurological comorbidities (encephalitis and meningitis).

Max Blood lactate levels as indicator of shock

Min Ferrtin Level

Coagulation Profile ( PT , APTT , TT , Platelet count , D-dimer , Plasma fibrinogen) Hypercoagulation (D-dimers, lupus anticoagulant, cardiolipin antibodies and anti-B2G1 antibodies)

# APPENDIX 1:​

Diagnostic codes for COVID-19 severity (provided by Lea Davis and Julia Sealock) <https://drive.google.com/file/d/1ck0ABYZ6oYnMStoYnGpnA7n1W6wcY3_6/view>

**Table S4.** Descriptive statistics of the Generation Scotland, Helix, Lifelines and NTR cohorts. Due to absence of testing data, Generation Scotland could not be used for replication of the Menni COVID-19 prediction model and in the development of the Lifelines COVID-19 prediction model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Helix** | **Lifelines** | **NTR** |
| Positive test  | Negative test | Positive test  | Negative test | Positive test | Negative test |
| N (%) | 27 (12.5) | 189 (87.5) | 56 (8.7) | 586 (91.3) | 85 (18.2) | 382 (81.8) |
| Age, yr, mean (sd) | 49 (14.4) | 51 (15.3) | 52.4 (9.2) | 52.2 (12.3) | 41.0 (13.1) | 41.3 (12.9) |
| Male sex, N (%) | 11 (40.7) | 58 (30.7) | 14 (25.0) | 155 (26.5) | 10 (13.3) | 65 (86.7) |
| BMI, kg/m2, mean (sd) | 28.6 (7.8) | 28.7 (6.4) | 26.2 (4.3) | 26.4 (4.7) | 25.4 (4.6) | 25.4 (4.8) |
| Current smoker, N (%) | 1 (4.0) | 18 (10.0) | 3 (5.4) | 52 (8.9) | 6 (11) | 48 (89) |

**Table S5.** Reported number of individuals with an essential occupation and who have been in contact with infected individuals across negative, positive, and predicted COVID-19 cases in the Lifelines cohort.

|  |  |  |
| --- | --- | --- |
| **LifelinesCOVID-19 cohort** | **Works in an essential occupation (%)** | **Been in contact with an individual with a confirmed SARS-CoV-2 infection (%)** |
| Negative test | 358 (61.1) | 113 (19.3) |
| Positive test | 43 (76.8) | 35 (62.5) |

**Table S6**. Overlapping symptoms in the Helix, Lifelines and NTR cohorts. Logistic regression for each symptom separately in Lifelines on positive (n=56) vs negative (n=586) tested subjects to define symptom cut-offs (reference = absence of symptom).

|  |  |  |  |
| --- | --- | --- | --- |
|   | **Low symptom severity** | **Intermediate/high symptom severity** | **Cut-off to define positive symptom** |
|   | OR (p) | OR (p) |   |
| Difficulty breathing | 2.57 (0.003) | 2.82 (0.006) | Low |
| Runny nose | 1.01 (0.959) | 2.68 (0.006) | Intermediate/high |
| Sore throat | 1.88 (0.064) | 2.42 (0.014) | Low |
| Cough | 4.31 (0.009) | 9.29 (0.000) | Low |
| Fever | 11.83 (0.000) | 14.66 (0.000) | Low |
| Diarrhea/stomach ache | 1.87 (0.048) | 2.43 (0.018) | Low |
| Loss taste/smell | 2.47 (0.047) | 21.03 (0.000) | Intermediate/high |
| Fatigue | 1.85 (0.306) | 7.22 (0.000) | Intermediate/high |

**Table S7a**. The Lifelines COVID-19 prediction model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **B** | **SE** | **p** | **OR** | **OR 95% CI** |
| Constant | -4.497 | 0.541 | 0.000 | 0.011 |  |
| Cough | 1.032 | 0.560 | 0.065 | 2.807 | 0.938 - 8.406 |
| Fever | 2.042 | 0.343 | 0.000 | 7.709 | 3.934 - 15.106 |
| Loss of smell or taste | 2.145 | 0.363 | 0.000 | 8.545 | 4.197 - 17.397 |

**Table S7b**. Diagnostics of different cut-offs of predicted probability of the Lifelines COVID-19 prediction model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cut-off predicted probability** | **Sensitivity** | **Specificity** | **Positive predicted value** | **Negative predicted value** |
| > 0.10 | 0.821 | 0.848 | 0.341 | 0.980 |
| > 0.20 | 0.500 | 0.952 | 0.500 | 0.952 |
| > 0.30 | 0.357 | 0.974 | 0.571 | 0.941 |
| > 0.40 \* | 0.357 | 0.974 | 0.571 | 0.941 |
| > 0.50 | 0.339 | 0.974 | 0.559 | 0.939 |

\* A predicted probability > 0.4 was used to define a positive predicted case.

**Table S7c**. Model diagnostics of the Lifelines COVID-19 prediction model in the Helix, Lifelines and NTR cohorts.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Cohort** | **AUC (95% CI)** | **Sensitivity** | **Specificity** | **Positive predicted value** | **Negative predicted value** |
| Helix | 0.742 (0.686-0.869) | 0.333 | 0.943 | 0.444 | 0.912 |
| Lifelines | 0.882 (0.834-0.930) | 0.357 | 0.974 | 0.571 | 0.941 |
| NTR | 0.848 (0.801-0.895) | 0.422 | 0.958 | 0.700 | 0.878 |

The model: -4.497 + 1.032\*cough + 2.042\*fever + 2.145\*loss of smell or taste. A predicted probability cut-off of > 0.40 is used to define a positive predicted case.

**Table S8**. Comparison of the case predictions of the Lifelines COVID-19 prediction model and the Menni COVID-19 prediction model in the Helix, Lifelines and NTR cohorts.

|  |  |  |
| --- | --- | --- |
|  | **Tested negative** | **Tested positive** |
| **Menni *et al.*** **model****Lifelines** **model** | Predicted negativeN (%) | Predicted positiveN (%) | Predicted negativeN (%) | Predicted positiveN (%) |
| Cohort: Helix |
| Predicted negative | 471 (93.8) | 19 (3.8) | 29 (60.4) | 4 (8.3) |
| Predicted positive | 1 (0.2) | 11 (2.2) | 0 (0.0) | 15 (31.3) |
| Cohort: Lifelines |
| Predicted negative | 557 (95.1) | 14 (2.4) | 30 (53.6) | 6 (10.7) |
| Predicted positive | 0 (0.0) | 15 (2.6) | 1 (1.8) | 19 (33.9) |
| Cohort: NTR |
| Predicted negative | 325 (90.8) | 18 (5.0) | 34 (41.5) | 13 (15.9) |
| Predicted positive | 10 (2.8) | 5 (1.4) | 14 (17.1) | 21 (25.6) |

**Table S9.** Downstream analysis of DEPICT