**UKB COVID GxE Interaction Analysis**

**“Gene-environment interaction analysis incorporating sex, cardiometabolic diseases, and multiple deprivation index reveals novel genetic associations with COVID-19 severity”**

**Introduction**

Epidemiological research has uncovered multiple risk factors for COVID-19 severity, including sex, metabolic conditions (such as type 2 diabetes and obesity), and socioeconomic status. Male sex is independently associated with higher mortality and worse COVID-19 outcomes (Palaiodimos et al., 2020, Park et al., 2020). Cardiometabolic conditions, such as Type 2 diabetes (T2D) and obesity, are also associated with increased COVID-19 susceptibility and severity (Barron, 2020; Zhu et al., 2020). Additionally, associated comorbidities of obesity, such as deregulated immune response, chronic inflammation, metabolic dysfunction, and compromised cilia on airway epithelial cells may put individuals at higher risk of severe COVID-19 (Ritter et al., 2020). Minoritized communities are disproportionately impacted by COVID-19 and may be predisposed to worse conditions due to environmental factors, limited healthcare access, and other societal factors (Tai et al., 2020). Furthermore, housing and neighborhood density and increased work-related exposure may put low-income groups at higher risk (Burström & Tao, 2020). Additionally, the greater prevalence of underlying chronic conditions among individuals with lower socioeconomic status puts this group at greater risk of severe outcomes.

Genetic investigations, such as that from the Host Genetics Initiative (HGI) consortium, have demonstrated that specific genomic regions are associated with COVID-19 severity. The HGI global meta-analysis identified 13 genome-wide significant loci, 9 of which were associated with increased risk of severe symptoms for hospitalized COVID (Ganna, 2021). Several loci were further associated with interstitial lung disease and autoimmune and inflammatory diseases, possibly predisposing individuals to greater immune response and worse outcomes.

It is not clear whether genetic factors impact the relationship between these key risk factors and COVID-19 severity, or whether these interactions can uncover novel genetic loci impacting this outcome. We sought to understand the interactions between genetic variants and risk factors, to gain novel understanding of the underlying mechanisms impacting COVID-19 severity, and add an important dimension to the current epidemiological literature on COVID-19. We undertook a series of three genome-wide gene-environment interaction studies in the UK Biobank, while conducting both interaction effect tests and joint tests of genetic main and interaction effects. The “environmental” exposures included sex, cardiometabolic health (obesity and type 2 diabetes status), and social determinants of health (as quantified by the multiple deprivation index). The binary outcome was severe COVID-19 (as defined by hospitalization or death due to COVID-19) while the rest of the population was used as a control group. Using GxE analyses and GWAS post-processing methods, we found 5 genome-wide significant loci that provide insight into the biological mechanisms of severe COVID-19 outcomes.

**Methods**

*UK Biobank Dataset*

The UK Biobank (UKB) is a population-based cohort including over 500,000 individuals living in England, Wales, and Scotland. The sub-population of interest for this study included unrelated individuals of European ancestry in order to minimize genetic heterogeneity. Sample sizes varied depending on available phenotypes across these populations. COVID-19 test results were downloaded from the UKBB data portal on January 1, 2020. The severe COVID-19 phenotype was defined as laboratory confirmed SARS-CoV-2 infection plus hospitalized COVID-19, with the rest of the population serving as controls. This definition was designed to mirror that of the “B2” phenotype used by the COVID-19 Host Genetics Initiative team (COVID-19 Host Genetics Initiative, 2020) and is outlined in Supp. Fig. 1. Genotype preprocessing was primarily performed by the UKB with filters at the marker and sample level. At the sample level, high heterozygosity and over 5% missingness was reported. Genotypes were further subsetted to common variants (minor allele frequency > 0.05) for analysis.

*Exposures of Interest*

Risk factors used as exposures were measures of genetically-determined sex, cardiometabolic health, and social determinants of health (SDH). For cardiometabolic measures, BMI was used as a measure of obesity and T2D status was determined based on self-reported medical history and medication use (“probable” or “possible” algorithmic definitions described by Eastwood an colleagues ()). BMI and T2D were tested jointly, and then individually as a sensitivity analysis. The multiple deprivation index (MDI) was used as a measure of social determinants of health (SDH). The MDI is composed of metrics including economic stability, physical environment, and education; details can be found at https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=76.

*Statistical analysis*

A genome-wide scan was performed based on a logistic regression model including gene-environment interaction terms*:*

Y was the binary severe COVID-19 indicator (defined above). The three genome-wide scans used the following exposures: sex, cardiometabolic conditions (BMI and T2D), and multiple deprivation index. For the cardiometabolic conditions, two environmental terms and two interaction terms were tested jointly. To test T2D exposure effect, GxT2D interaction obese and non-obese stratified analyses were run. Covariates included age, five genetic principal components, and sex. Genome-wide analysis was conducted using GEM v1.2 (CITE biorXiv) with robust standard errors. For each variant, two statistical tests were derived: an interaction test and a joint test of the interaction term(s) plus the genetic main effect.

Interaction and joint analyses were conducted on the Terra cloud platform. Phenotype definitions and population summaries were created in interactive Jupyter notebooks with an R 3.6 kernel. GWIS analyses were submitted as workflows using a Workflow Description Language (WDL) script implementing GEM. Post-GWIS summarization and visualizations were created in a separate Jupyter notebook. These notebooks can be viewed on GitHub (<https://github.com/manning-lab/ukb-covid-gxe>).

*Variant Biology Investigation*

Top variants were further investigated for trait associations, eQTLs, and linkage disequilibrium using dbSNP (NCBI), PhenoScanner (University of Cambridge, Staley, et al. 2016, Kamat, et al. 2019), RegulomeDB (Boyle et al., 2012), Type 2 Diabetes Knowledge portal (https://t2d.hugeamp.org/), and the LDlink tool (Myers, Machiela, & Chanock, 2020). Colocalization between interactions and eQTLs was performed using the *coloc* package (CITE: <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004383>) along with blood-based eQTL summary statistics from the eQTLGen Consortium (CITE: <https://doi.org/10.1101/447367>).

**Results**

The UKB population is described in Table 1a, with subjects categorized into those having experienced severe COVID-19 (hospitalization or death from COVID-19; see Methods) and the remaining population (regardless of infection status). The overall population had a greater proportion of females, while cases were more likely to be male (p=1.11e-16, OR [95% CI] =1.41[1.30-1.52]). Cases also had a greater prevalence of T2D (OR=2.57[2.25-2.93]) and higher values of BMI (OR=1.92[1.77-2.09]) and MDI (OR=1.36[1.32-1.41]) as shown in Table 1b.

We conducted a GWIS for each of the following exposures: sex, cardiometabolic traits (BMI and T2D, tested jointly), and MDI. Top index variants after pruning are displayed in Suppl. Tables S2-4. Across all scans, five variants (rs2268616, rs182113773, rs148793499, rs11115199, and chr2:218260234) passed a genome-wide significance (GWS) threshold (p < 5e-08) in the joint test. One of these five (rs11115199) was additionally found to be GWS in the cardiometabolic (CM) interaction test (Figure 2; Table 2). Two of these variants (rs148793499, rs11115199) passed a study-wide significance threshold (p < 5e-08 / 3 exposures = 1.6e-8). Of the five variants, a GWS marginal effect was identified for only rs2268616 (p=1.08e-08) and rs182113773 (p=1.39e-08). This result demonstrates that the joint test discovered variants that would not have been found via a standard GWAS in this population.

These five GWS variants were compared to genetic main effects from the HGI meta-analysis (with UKB omitted) testing the equivalent “B2” phenotype (hospitalized COVID-19 vs. population). A significant genetic main effect would constitute a partial replication of the joint test (genetic plus interaction effect) hypothesis. Neither of the two variants available in the HGI summary statistics showed nominal replication (both p>0.05). For the remaining three, neither the variants nor reasonable genetic proxies (r2 > 0.5 using European-based linkage disequilibrium patterns) were available in the HGI dataset.

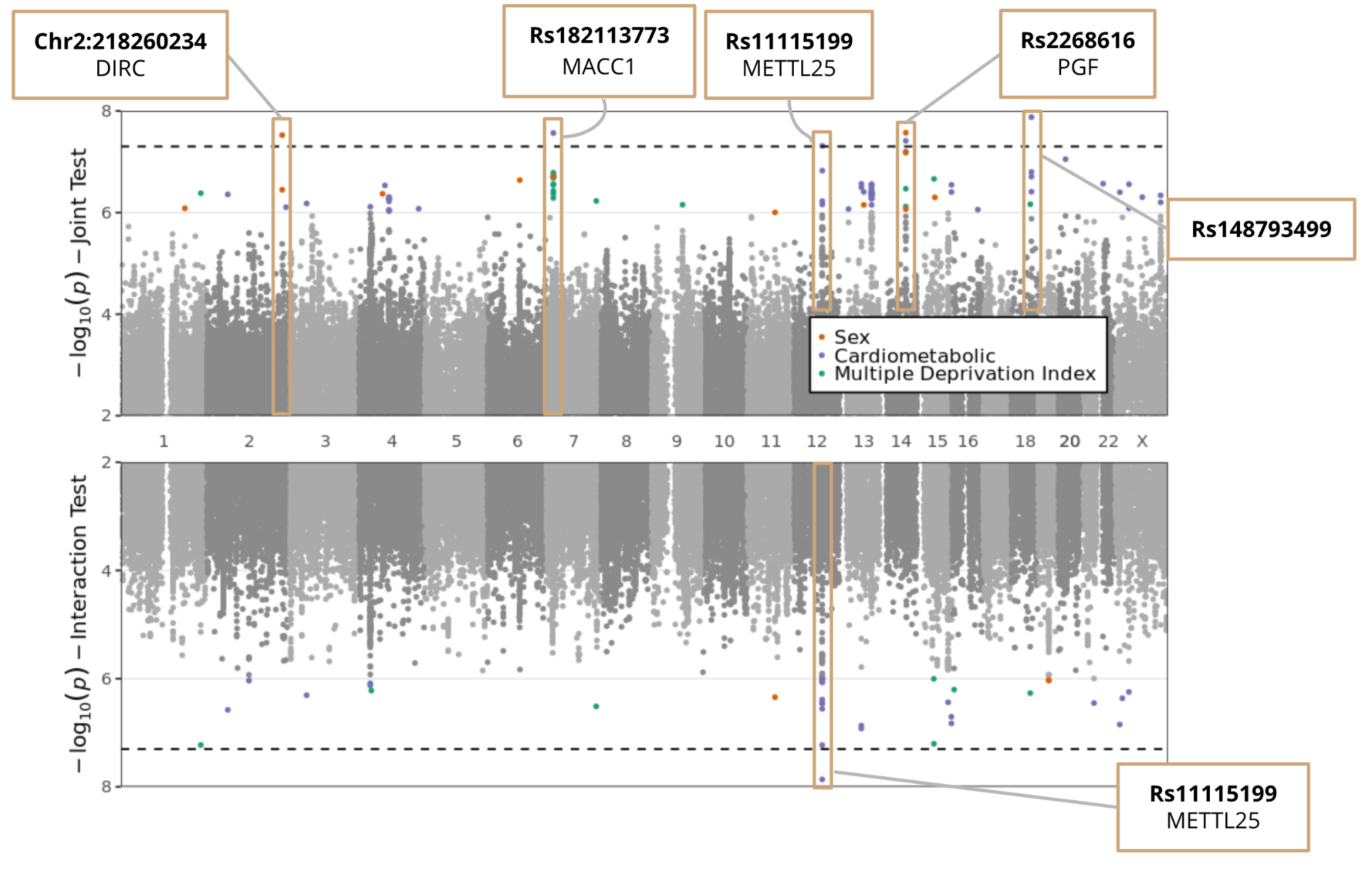
Next, we further explored these top variants and interactions to understand their potential biological function. One variant of interest, rs2268616 (MAF=0.018), was genome-wide significant in the joint sex analyses (p=2.67e-08) and joint cardiometabolic diseases (p=3.87e-08). This variant sits in an intron of the placental growth factor (PGF) gene and is associated with testosterone in GWAS analyses. It is also a putative enhancer in lung and other tissues, and is an eQTL for EIF2B2 (a gene in a family of proteins that regulate viral mRNA translation) in whole blood. However, colocalization analysis using whole blood eQTL statistics from eQTLGen Consortium did not support the hypothesis of a shared causal variant with either PGF or EIF2B (posterior probabilities <0.1%). Sex-stratified analysis showed a stronger genetic effect in males (OR [95% CI] =1.79 [1.43-2.24]) compared to females (OR=1.45 [1.11-1.9]) as shown in Figure 3A. Stratified T2D results also showed a greater genetic effect in T2D cases (OR=2.01 [1.22-3.32]) compared to no T2D cases (OR=1.6 [1.33-1.02]). Using the LocalZoom software, the rs2268616 was compared with HGI results for genetic main effects on the same COVID-19 severity phenotype (Figure 3B). Within our dataset, we noticed nearby signals to our variant of interest, rs2268616. Furthermore, rs2268616 is in linkage disequilibrium (p<0.001) with a significant variant from the HGI analysis (rs2005863, p\_HGI=0.001, MAF=0.0307), providing support for the effect of this variant on COVID-19 severity.

Our other variants also indicated genetic effects on COVID-19 severity mediated through interaction effects. Rs182113773 was genome-wide significant for the cardiometabolic joint test (p=2.71e-08) and found in the intron for MACC1. This variant sits in an enhancer within neutrophils, monocytes, and B cells and has a RegulomeDB score of 0.59, suggesting a regulatory role in MACC1 transcription. Variant chr2:218260234 was genome-wide significant in the sex analysis joint test (p=2.99e-08, MAF=0.025). Stratified analysis demonstrated a strong genetic effect in males (OR=1.8 [1.47-2.19]) that was not found in females (OR=1.03 [0.779-1.35]). In addition, rs11115199 was a genome-wide significant intergenic variant for the cardiometabolic interaction and joint test (respectively, p=1.37e-08 & 4.85e-08, MAF=0.02). eQTL relationships from GTEx show expression of METTL25. A proximal variant, rs79719968 (R2=0.176, D’=0.849), was also associated with METTL25 and the lung tissue. Similarly, rs148793499 is a genome-wide significant intergenic variant for the cardiometabolic joint test (p=1.8e-10, MAF=0.01). Stratified genetic effects from logistic regression showed pronounced associations strongly in obesity (OR=2.36 [1.7-3.27]) and slightly in T2D (OR=2.01 [1.03-3.93]).

Our social determinants of health analysis did not identify any genome-wide significant variants. However, rs2268616 was suggestively significant in the joint test (p=3.4E-07). Stratified tests showed higher quartiles of MDI had greater genetic effects in 4 out of the top 15 variants (Suppl. Fig. 4).

**Table 1: Characteristics of European ancestry samples from the UK Biobank cohort.** We present **the mean and standard deviation for continuous covariates, percentage of the sample for dichotomous covariates and odds ratio and P value for association with Severe COVID from a multivariate logistic regression model.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 1. Population characteristics stratified by COVID severity**  (Total N=378,051) | | | |  |  |
|  | Overall | Control  (N=375,649) | Severe COVID  (N=2,402) | P value | Odds Ratio (95% Confidence Interval) |
| **Age (years)** | 56.73 + 8.02 | 56.7 + 8 | 57.9 + 8.6 | YY | ZZ |
| **Sex (Male)** | 46% | 46% | 54% | 1.11e-16 | 1.41 [1.30-1.52] |
| **Body Mass Index (kg/m2)** | 27.37 + 4.76 | 27.4 (4.8) | 29.3 (5.4) | 1.41e-53 | 1.92 [1.77-2.09] |
| **Type 2 Diabetes** | 4% | 4% | 10% | 7.24e-45 | 2.57 [2.25-2.93] |
| **Multiple Deprivation Index** | 16.88 + 13.5 | 16.8 (13.5) | 22.1 (16.5) | 5.50e-6 | 1.36 [1.32-1.41] |



**Figure 1: Plots of sex, cardiometabolic, and MDI joint and interaction tests.** Theupper plot shows negative logarithm of joint p-values in a test of main and interaction effects, while the lower plot shows negative logarithm of the interaction test p-values. X-axis corresponds to genomic position. Genome-wide significant results are labeled with the most significant variant at the locus and the nearest gene within XX MB.

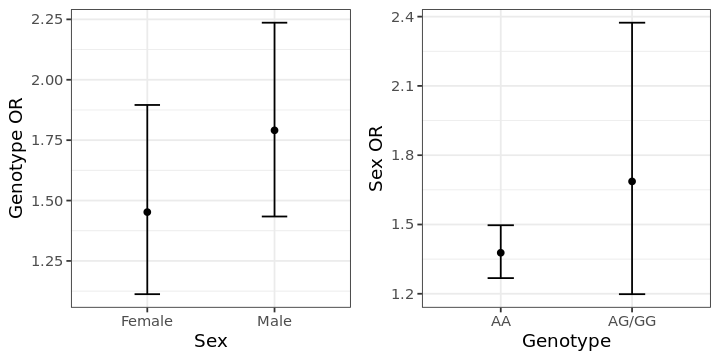
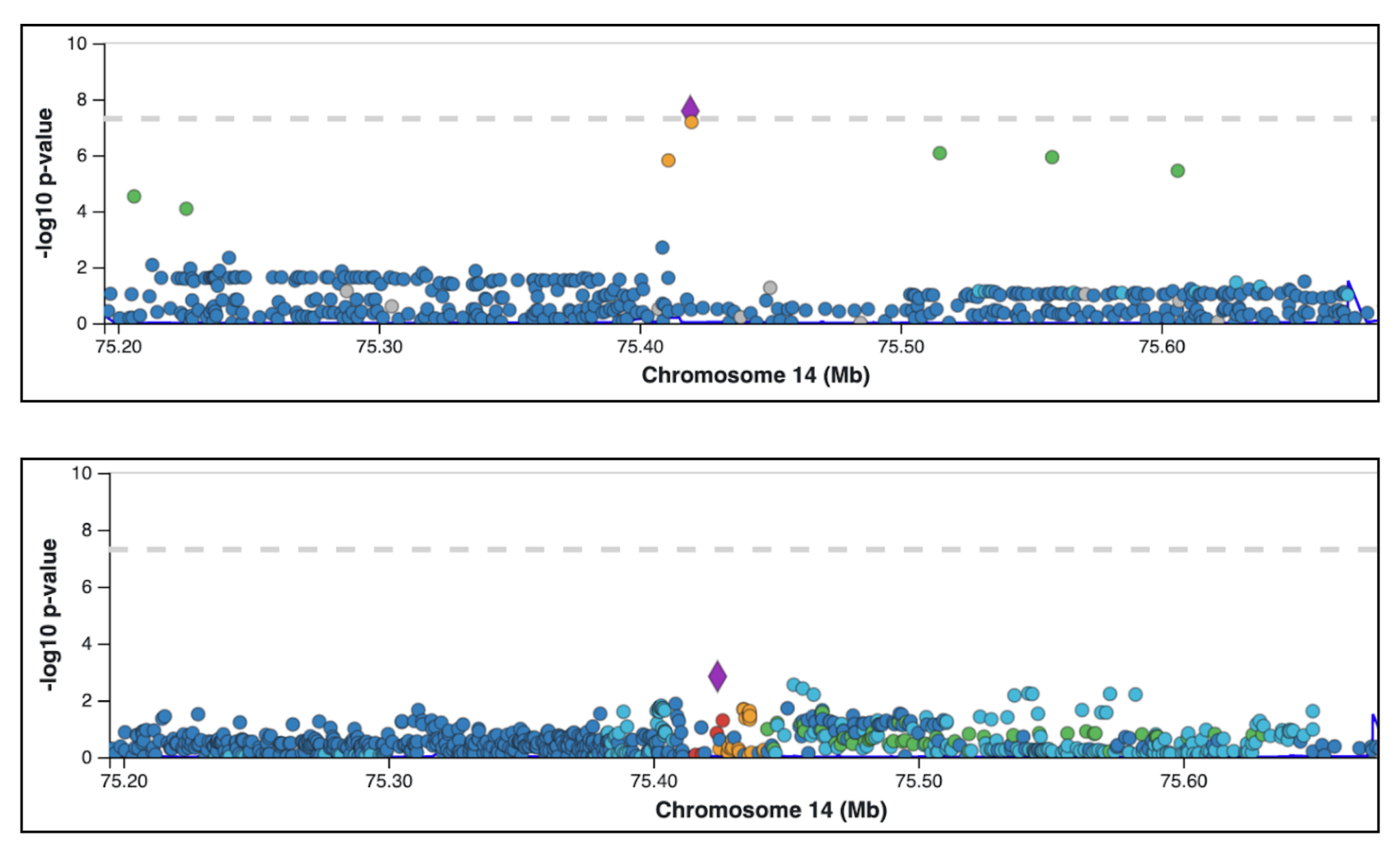
**Table 2: Genome-wide significant associations from interaction and joint tests. A**. Sex interaction and joint tests. **B.** Cardiometabolic interaction and joint tests.

A.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RSID** | **location** | **Effect Allele** | **Non-Effect Allele** | **Eff\_Allele\_Freq** | **Interaction p-value** | **Joint p-value** | **OR interaction** | **OR combined** | **OR in males** | **OR in females** |
| rs2268616 | 14:75419444 | G | A | 0.018 | 0.14 | 2.7E-08 | 1.2 [0.87-1.7] | 1.6 [1.4-1.9] | 1.8 [1.4-2.2] | 1.4 [1.1-1.9] |
| 2:218260234\_AC\_A | 2:218260234 | A | AC | 0.026 | 0.00013 | 3.0E-08 | 1.7 [1.2-2.4] | 1.4 [1.2-1.7] | 1.8 [1.5-2.2] | 1.0 [0.78-1.3] |

B.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RSID** | **location** | **Effect Allele** | **Non-Effect Allele** | **Eff\_Allele\_Freq** | **Interaction p-value** | **Joint p-value** | **OR combined** | **OR in no\_t2d** | **OR in T2D** | **OR in no\_obesity** | **OR in obesity** |
| rs148793499 | 18:58314588 | C | T | 0.010 | 8.4E-06 | 1.3E-08 | 1.6 [1.3-2.03] | 1.6 [1.2-2.0] | 2.0[1.0-3.9] | 1.2[0.83-1.7] | 2.4[1.7-3.3] |
| rs11115199 | 12:82510665 | T | G | 0.020 | 1.4E-08 | 4.8E-08 | 0.91 [0.74-1.1] | 0.72 [0.56-0.92] | 2.6 [1.7-3.9] | 1.0 [0.74-1.3] | 0.85 [0.59-1.2] |
| rs182113773 | 7:20239837 | A | C | 0.015 | 0.053 | 2.7E-08 | 1.7  [1.4-2.1] | 1.6 [1.3-2.0] | 2.6 [1.6-4.3] | 1.6[1.3-2.1] | 1.9[1.4-2.6] |
| rs2268616 | 14:75419444 | G | A | 0.018 | 0.26 | 3.9E-08 | 1.6 [1.4-1.9] | 1.6 [1.3-1.9] | 2.0 [1.2-3.3] | 1.8[1.5-2.2] | 1.4 [1.0-1.9] |

1. ****
2. 

**Figure 3:** Inspection of the sex-rs2268616 interaction effect. (A) Sex-stratified genetic effects on severe COVID-19 after adjustment for the primary set of covariates. Y*-*axis indicates the estimated odds ratio for severe COVID-19 per alternate allele. (B) Regional association plots showing association signals from this analysis (sex joint test, top panel) and HGI B2 phenotype meta-analysis (genetic main effects with UKB omitted, bottom panel).

**Discussion**

Exploring the interplay of genetics and sex offers novel understanding of the underlying mechanisms impacting COVID-19 severity and adds an important dimension to the current epidemiological literature on COVID-19. In this genome-wide gene-environment interaction analysis, we found five significant genomic regions (p<5e-8) that interact with well-established risk factors to influence COVID-19 severity.

Sex-dimorphic transcripts and hormones, as well as differences in environmental factors between the sexes, contribute to differential immune responses between sexes and may mediate the established association of male sex with greater COVID-19 severity (Klein & Flanagan, 2016). In our analysis, rs2268616 was statistically significant in the joint analyses for sex and cardiometabolic diseases (p<5e-8). This variant is associated with testosterone and placental growth factor gene in GWAS analyses, suggesting that this variant interacts with sex to mediate worse COVID-19 outcomes. Interestingly, this variant also has an eQTL for *EIF2B2*, a gene within a family of proteins that mediate viral mRNA translation. This supports previous work that has shown that the coronavirus can enter the testis, delaying viral RNA clearance in males (Shastri et al., 2020). Moreover, a study found an increased odds ratio of dying and significantly increased levels of inflammatory markers in male COVID-19 positive hospitalized patients compared with women (Lau et al., 2021). The *EIF2B2* variant is linked to a strong transcription chromatin state in the cells of the lung, spleen, and B-cells, perhaps mediating the robust inflammatory response in males that is associated with worse COVID outcomes. Furthermore, rs2268616 sits within an enhancer in lung tissue, suggesting a role of this variant on transcription and respiratory complications after SARS-CoV-2 infection. rs2268616 has also been shown to have positive associations with height and coronary artery disease and a negative association with HOMA-B, representing a potential broad influence on metabolic traits. Our findings suggest that this genetic variant may modify the relationship between biological differences and associated worse COVID-19 outcomes primarily through regulating viral RNA clearance immune response and lung cell transcription.

Comorbidities associated with cardiometabolic health such as obesity and T2D have been implicated in mediating worse COVID-19 outcomes (Ritter et al., 2020). Our findings show four variants that were genome-wide significant in our cardiometabolic joint tests: rs182113773, rs11115199, rs148793499 and rs2268616. Located within the intron for *MACC1*, a gene associated with BMI-adjusted waist circumference and BMI-adjusted waist-hip ratio, rs182113773 is an enhancer within neutrophils, monocytes, and B cells. This variant also has high gene expression in EBV-transformed lymphocytes and is a likely regulatory variant (RegulomeDB score of 0.59), which further suggests that the interaction of this variant with cardiometabolic health has a regulatory role on immune response. Studies found that increased neutrophil count in T2D groups are associated with clinical severity and may mediate the positive association between T2D and COVID-19 severity (Zhao et al., 2020). Thus, this *MACC1* variant may be interacting with cardiometabolic health to mediate greater COVID-19 severity. Furthermore, obese adipose tissues overexpress receptors and proteases that enable the entry of SARS-CoV-2, possibly contributing to the severe inflammation and immune response of individuals with obesity (Ritter et al., 2020).

The rs11115199 variant is an eQTL for *METTL25*, which is associated with BMI, and shows low transcription in memory T cells and B cells, indicating an interaction with cardiometabolic health that may modify immune response. T2D knowledge portal also shows positive associations with weight and T2D adjusted by BMI, negative association with obesity, indicating this variant may predispose individuals to worse COVID-19 outcomes by interacting with cardiometabolic health and associated comorbidities. In the cardiometabolic test, rs148793499 had low transcription in CD14-positive monocytes and lung fibroblasts. In stratified cardiometabolic tests, the variant showed slight genetic effects in individuals with T2D and strong genetic effects in individuals with obesity.

Alongside decreased immune response mediated by testosterone, rs2268616 may also play a role in the deflated immune response seen in cases with cardiometabolic disease status. This variant has a positive association with coronary artery disease and a negative association with HOMA-B (a method that assesses β-cell function from basal fasting glucose and insulin). For cardiometabolic diseases, well controlled blood glucose and smaller glycemic variability have been associated with lower mortality during hospitalization due to COVID-19. Therefore, this variant may help explain the COVID-19 biology that increases the risk for individuals with T2D. Interactions between these genetic factors and deregulated immune response, chronic inflammation, metabolic dysfunction, and other comorbidities of obesity and T2D may be placing individuals at greater risk for worse outcomes of COVID-19.

Our analysis focusing on social determinants of health did not identify any significant variants. This may be a function of the noise associated with the MDI measurement and the difficulty in using this measurement to represent social determinants of health in a large diverse population. One study leveraged an Index of Multiple Deprivation and Income Deprivation Affecting Older People Index to show higher incidence of COVID-19 related deaths in the most deprived quartiles (Bach-Mortensen & Degli Esposti, 2021).We subsetted our sample to participants from England to reduce heterogeneity, but this reduced the sample size used for the MDI analysis (by 16.5%; 2,007 vs. 2,402 cases) reduced statistical power. Another possibility for the lack of signal is that the effects of genetics and social determinants of health on COVID-19 severity are relatively independent and have no underlying interactions to be uncovered. In our stratified tests, higher quartiles of multiple deprivation index had greater genetic effects in 4 out of the top 15 variants (Suppl. Fig. 4). Thus, better measurements of social determinants of health and an increased population sample may uncover novel variants that have not yet been identified.

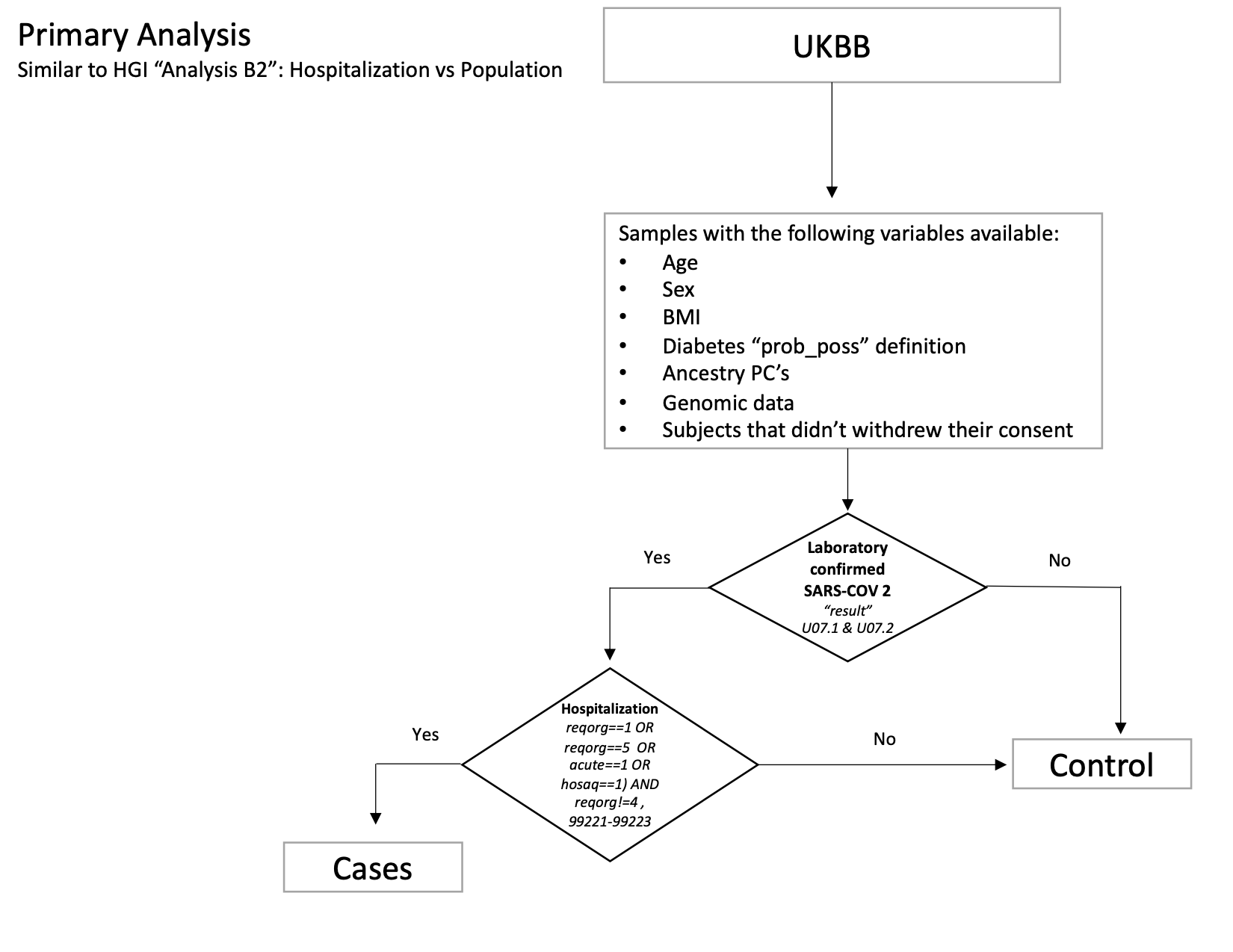
The results of this study may be limited due to linkage disequilibrium and heterogeneity caused by geographic location within our sample population. The case definition allows us to identify variants associated with severity, however these results need to be taken with precaution given the possibility of Collider bias. Analyzing UK Biobank data, the participants tested for COVID-19 were highly selected for a range of genetic, behavioral, cardiovascular, demographic, and anthropometric traits (Griffith, 2020).

By subsetting our dataset to European ancestry, we reduce the heterogeneity but face a limited sample size. Nonetheless, leveraging this GxE interaction analysis uncovers novel variants and highlights biological mechanisms of interest. The GEM marginal p-value did not pass the genome-wide significant threshold for three of the five variants, indicating that our GxE interaction methods were able to uncover genetic effects previously unidentified.

Our findings suggest that some GxE interaction effects may contribute to the differences in COVID-19 severity, demonstrating COVID-19 severity is influenced by both gene and environment. Sex-associated difference in immune response and cardiometabolic disease comorbidities that deregulate immune response may interact with the identified genetic variants and put individuals at higher risk for worse outcomes of COVID-19. Future studies investigating the stratified effects of sex, T2D and BMI, and social determinants of health on COVID-19 susceptibility, as well as similar analysis with a wider array of ancestries, may further reveal underlying the genetic interaction effects that place individuals at higher risk.

**Supplemental Figures**

**Supp. Fig. S1**: Hospitalization vs Population Phenotype Flow Chart



**Supp. Table 1:** Population characteristics of European ancestry samples extracted from UKB, stratified by sex.

|  |  |  |
| --- | --- | --- |
| **Supp. Table 1. Population characteristics stratified by sex** (n=378,051) | | |
| **Sex** | Female  (n= 203,961) | Male  (n= 174,090) |
| **Age** | 56.5 (7.9) | 57 (8.1) |
| **BMI** | 27 (5.1) | 27.8 (4.2) |
| **T2D** | 3% | 6% |
| **MDI** | 16.6 (13.2) | 17.2 (13.8) |

**Supp. Table 2:** Top variants from the sex analysis (joint and interaction tests) along with stratified genetic odds ratios.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RSID** | **location** | **Effect Allele** | **Non-Effect Allele** | **Effect Allele**  **Freq** | **GEM Interaction p-value** | **GEM Joint p-value** | **OR interaction** | **OR combined** | **OR (male)** | **OR (female)** |
| rs2268616 | 14:75419444 | G | A | 0.018 | 0.139689 | 2.67E-08 | 1.23 [0.868-1.75] | 1.64 [1.38-1.94] | 1.79 [1.43-2.24] | 1.45 [1.11-1.9] |
| 2:218260234\_AC\_A | 2:218260234 | A | AC | 0.026 | 0.000128 | 2.99E-08 | 1.74 [1.24-2.45] | 1.44 [1.22-1.69] | 1.8 [1.47-2.19] | 1.03 [0.779-1.35] |
| rs182113773 | 7:20239837 | A | C | 0.015 | 0.866041 | 1.96E-07 | 0.98 [0.657-1.45] | 1.73 [1.42-2.1] | 1.71 [1.3-2.23] | 1.75 [1.31-2.34] |
| rs117993077 | 6:93824151 | T | G | 0.012 | 8.86E-06 | 2.3E-07 | 2.67 [1.55-4.57] | 1.47 [1.16-1.85] | 2.06 [1.57-2.7] | 0.78 [0.487-1.24] |
| rs55926550 | 4:68007466 | C | G | 0.012 | 0.002296 | 4.27E-07 | 0.59 [0.388-0.911] | 1.6 [1.3-1.98] | 1.22 [0.882-1.69] | 2.05 [1.55-2.7] |
| rs11232355 | 11:80540380 | G | A | 0.069 | 4.54E-07 | 9.92E-07 | 0.58 [0.463-0.725] | 1.09 [0.972-1.21] | 0.82 [0.694-0.972] | 1.42 [1.22-1.64] |
| rs117234438 | 15:52349607 | G | A | 0.011 | 0.000927 | 5.01E-07 | 1.93 [1.17-3.19] | 1.64 [1.3-2.06] | 2.09 [1.59-2.76] | 1.08 [0.71-1.64] |
| rs796925608 | 13:68788267 | ACTTTGAAAATAAC | A | 0.898 | 0.841366 | 7.06E-07 | 1.02 [0.827-1.26] | 0.76 [0.681-0.839] | 0.76 [0.662-0.879] | 0.75 [0.64-0.872] |
| rs112561874 | 1:185427460 | A | G | 0.016 | 3.54E-05 | 8.24E-07 | 0.47 [0.305-0.723] | 1.42 [1.15-1.74] | 0.94 [0.668-1.32] | 2 [1.53-2.6] |
| rs12461506 | 19:30878810 | C | G | 0.643 | ﻿ 9.24304e-07 | ﻿5.89398e-06 | 1.35 [1.19-1.52] | 1 [0.943-1.06] | 1.15 [1.06-1.25] | 0.85 [0.783-0.932] |
| rs17066139 | 3:62113696 | T | C | 0.115 | 0.013713 | 1.17E-06 | 1.22 [1.03-1.44] | 1.22 [1.12-1.32] | 1.33 [1.19-1.48] | 1.09 [0.955-1.24] |
| rs114103616 | 21:16998740 | T | A | 0.028 | 0.014151 | 1.53E-06 | 0.7 [0.501-0.983] | 1.47 [1.24-1.74] | 1.23 [0.964-1.58] | 1.77 [1.4-2.22] |
| rs182465934 | 1:22335633 | A | C | 0.028 | 4.16E-05 | 1.88E-06 | 1.93 [1.35-2.77] | 1.3 [1.1-1.54] | 1.67 [1.36-2.05] | 0.87 [0.645-1.17] |
| rs758053125 | 20:6586718 | T | C | 0.050 | 8.37E-06 | 2E-06 | 0.52 [0.379-0.712] | 1.22 [1.04-1.43] | 0.87 [0.688-1.11] | 1.68 [1.36-2.06] |
| rs114807731 | 3:60122415 | A | G | 0.022 | ﻿2.98353e-05 | ﻿2.10734e-06 | 1.99 [1.37-2.9] | 1.29 [1.09-1.53] | 1.68 [1.36-2.07] | 0.84 [0.616-1.15] |

**Supp. Table 3:** Top variants from the cardiometabolic (T2D and BMI) analysis (joint and interaction tests) along with stratified genetic odds ratios.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RSID** | **location** | **Effect Allele** | **Non-**  **Effect Allele** | **Effect Allele**  **Freq** | **Interaction p-value** | **Joint p-value** | **OR combined** | **OR (No T2D)** | **OR (T2D)** | **OR (No obesity)** | **OR (obesity)** |
| rs148793499 | 18:58314588 | C | T | 0.0102753 | 8.38293E-06 | 1.32208E-08 | 1.61 [1.27-2.03] | 1.56 [1.21-2.01] | 2.01 [1.03-3.93] | 1.18 [0.835-1.67] | 2.36 [1.7-3.27] |
| rs11115199 | 12:82510665 | T | G | 0.0204524 | 1.37218E-08 | 4.85379E-08 | 0.91 [0.736-1.12] | 0.72 [0.561-0.924] | 2.6 [1.73-3.91] | 0.96 [0.742-1.25] | 0.85 [0.592-1.22] |
| rs182113773 | 7:20239837 | A | C | 0.0147751 | 0.0529514 | 2.71316E-08 | 1.73 [1.42-2.1] | 1.62 [1.31-2.01] | 2.6 [1.57-4.3] | 1.63 [1.26-2.11] | 1.94 [1.43-2.64] |
| rs2268616 | 14:75419444 | G | A | 0.0179483 | 0.26394 | 3.87431E-08 | 1.64 [1.38-1.94] | 1.6 [1.33-1.92] | 2.01 [1.22-3.32] | 1.82 [1.48-2.24] | 1.38 [1.01-1.88] |
| rs6035512 | 20:19990781 | A | C | 0.0247661 | 0.000114499 | 8.88795E-08 | 1.41 [1.2-1.66] | 1.45 [1.22-1.72] | 1.15 [0.648-2.03] | 1.21 [0.97-1.52] | 1.74 [1.36-2.22] |
| rs141850011 | 13:61925813 | C | T | 0.0103374 | 1.20059E-07 | 2.72447E-07 | 1.17 [0.89-1.53] | 0.88 [0.639-1.22] | 3.67 [2.22-6.05] | 0.97 [0.669-1.41] | 1.53 [1.03-2.26] |
| rs2727176 | 15:100570438 | G | A | 0.0102964 | 1.48748E-07 | 2.84132E-07 | 1.21 [0.931-1.57] | 0.94 [0.686-1.28] | 3.79 [2.3-6.24] | 0.97 [0.669-1.4] | 1.56 [1.07-2.27] |
| rs6747163 | 2:60777763 | G | C | 0.178678 | ﻿ 2.6503e-07 | 4.39E-07 | 1.05 [0.977-1.14] | 0.99 [0.912-1.07] | 1.72 [1.39-2.12] | 1.11 [1.01-1.22] | 0.97 [0.858-1.11] |
| rs5747202 | 22:17969179 | G | T | 0.0374612 | 6.03274E-06 | 2.70082E-07 | 1.31 [1.1-1.56] | 1.15 [0.945-1.4] | 2.92 [1.93-4.42] | 1.26 [1-1.58] | 1.4 [1.05-1.85] |
| rs28864791 | 13:91502951 | C | T | 0.306404 | 5.06161E-05 | 2.78527E-07 | 0.89 [0.835-0.949] | 0.85 [0.792-0.908] | 1.32 [1.09-1.59] | 0.87 [0.805-0.948] | 0.91 [0.82-1.01] |
| rs144282550 | 4:74917179 | G | A | 0.0117938 | 0.00319523 | 2.91809E-07 | 1.65 [1.32-2.06] | 1.51 [1.19-1.92] | 2.97 [1.72-5.12] | 1.65 [1.25-2.18] | 1.7 [1.19-2.43] |
| rs180972330 | 21:40008094 | G | A | 0.0100758 | 3.54E-07 | 1.51E-06 | 1.03 [0.774-1.38] | 0.79 [0.555-1.11] | 3.72 [2.18-6.35] | 0.71 [0.454-1.1] | 1.64 [1.11-2.41] |
| rs113083888 | 15:91197671 | T | C | 0.0484105 | 3.65E-07 | 1.03E-06 | 1.06 [0.933-1.21] | 1.12 [0.977-1.28] | 0.6 [0.354-1.02] | 0.95 [0.793-1.13] | 1.27 [1.04-1.55] |
| rs796925608 | 13:68788267 | ACTTTGAAAATAAC | A | 0.897704 | 0.100633 | 3.93E-07 | 0.76 [0.681-0.839] | 0.77 [0.689-0.86] | 0.66 [0.487-0.902] | 0.75 [0.653-0.852] | 0.77 [0.647-0.912] |
| rs345361 | 4:86777776 | G | A | 0.969044 | 0.00084641 | 4.93E-07 | 0.73 [0.63-0.84] | 0.79 [0.672-0.922] | 0.43 [0.296-0.611] | 0.75 [0.624-0.905] | 0.71 [0.561-0.896] |

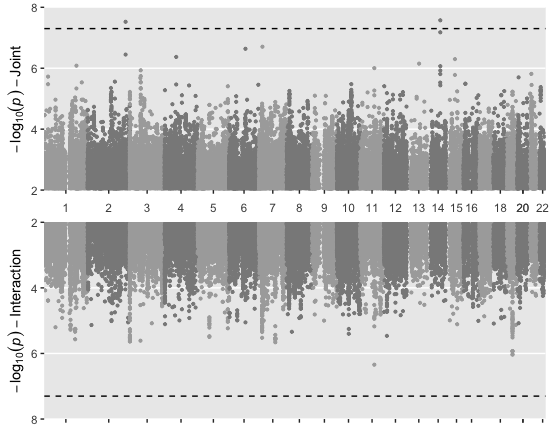
Obesity was defined as BMI ≥ 30 for stratification.

**Supp. Table 4:** Top variants from the multiple deprivation index analysis (joint and interaction tests) along with stratified genetic odds ratios.

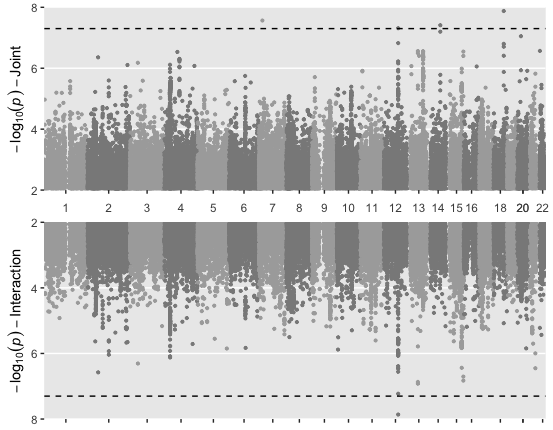
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **RSID** | **location** | **Effect Allele** | **Non-Effect Allele** | **Effect Allele Frequency** | **Interaction p-value** | **Joint p-value** | **OR interaction** | **OR combined** | **OR (Low MDI)** | **OR (High MDI)** |
| rs147319526 | 1:231731679 | G | A | 0.012486 | 5.93E-08 | 4.1588E-07 | 3.13 [1.41-6.91] | 0.98 [0.746-1.3] | 0.44 [0.212-0.893] | 1.35 [0.961-1.89] |
| rs76612061 | 15:49972935 | A | G | 0.026409 | 6.25E-08 | 2.1732E-07 | 1.95 [1.2-3.19] | 0.85 [0.697-1.03] | 0.56 [0.363-0.853] | 1.08 [0.848-1.37] |
| rs6966810 | 7:20141514 | A | G | 0.929889 | 0.00946 | 1.6439E-07 | 0.88 [0.694-1.1] | 0.8 [0.724-0.889] | 0.84 [0.692-1.01] | 0.72 [0.632-0.827] |
| rs73164732 | 7:144481768 | A | G | 0.011531 | 3.07E-07 | 5.914E-07 | 2.53 [1.26-5.09] | 1.17 [0.91-1.51] | 0.62 [0.332-1.17] | 1.6 [1.18-2.17] |
| rs2268616 | 14:75419444 | G | A | 0.018136 | 0.164618 | 3.3998E-07 | 0.8 [0.548-1.17] | 1.64 [1.38-1.94] | 1.89 [1.41-2.53] | 1.5 [1.17-1.92] |
| rs1646599 | 18:55236161 | A | G | 0.01178 | 5.4E-07 | 6.8523E-07 | 0.96 [0.841-1.1] | 1.05 [0.986-1.11] | 1.06 [0.948-1.17] | 1.02 [0.938-1.1] |
| rs139020188 | 4:35620323 | T | A | 0.010373 | 6.03E-07 | 1.6804E-06 | 2.55 [1.19-5.47] | 1.08 [0.81-1.45] | 0.62 [0.312-1.23] | 1.56 [1.11-2.18] |
| rs148817892 | 16:5934373 | G | C | 0.011508 | ﻿ 6.27546e-07 | ﻿ 2.51681e-06 | 4.25 [1.85-9.79] | 1.1 [0.85-1.43] | 0.38 [0.176-0.835] | 1.62 [1.19-2.18] |
| rs72752741 | 9:89168171 | T | C | 0.014311 | 0.000798 | 7.014E-07 | 1.08 [0.685-1.7] | 1.5 [1.23-1.84] | 1.45 [0.996-2.11] | 1.63 [1.25-2.13] |
| rs117488928 | 15:49563001 | C | A | 0.036988 | ﻿ 9.93601e-07 | ﻿ 5.7925e-06 | 1.79 [1.22-2.62] | 0.96 [0.822-1.12] | 0.69 [0.498-0.962] | 1.23 [1.02-1.5] |
| rs147349057 | 6:1059829 | TA | T | 0.015148 | 2.01E-06 | 1.2453E-06 | 2.43 [1.34-4.42] | 1.23 [0.98-1.54] | 0.69 [0.405-1.17] | 1.68 [1.28-2.22] |
| rs139833210 | 7:10216456 | A | T | 0.012896 | 0.001165 | 1.2661E-06 | 0.6 [0.377-0.954] | 1.41 [1.13-1.77] | 2.17 [1.55-3.03] | 1.29 [0.933-1.78] |
| rs192911167 | 2:89058008 | G | C | 0.010853 | ﻿1.57004e-06 | ﻿6.31539e-06 | 2.55 [1.21-5.38] | 1.22 [0.932-1.59] | 0.59 [0.301-1.14] | 1.5 [1.08-2.08] |
| rs113541905 | 15:59709733 | T | C | 0.043553 | ﻿ 1.72579e-06 | ﻿6.31187e-06 | 1.57 [1.12-2.18] | 1.04 [0.903-1.19] | 0.8 [0.602-1.06] | 1.25 [1.04-1.48] |
| rs11945368 | 4:37555712 | A | C | 0.17519 | ﻿ 0.0293873 | ﻿1.9631e-06 | 1.12 [0.948-1.33] | 1.18 [1.1-1.28] | 1.12 [0.979-1.29] | 1.26 [1.14-1.39] |

Low and high MDI were defined as being below or above the median value, respectively.

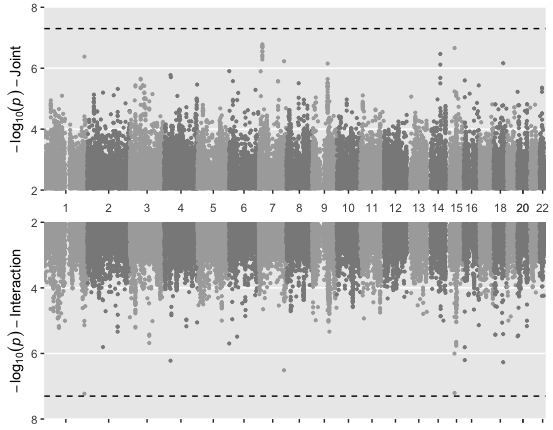
**Supp. Fig. S2**: Sex analysis joint and interaction analysis results. The Manhattan plot displays association test strengths for the joint (top panel) and interaction (bottom panel) tests as a function of genomic position (*x*-axis).

****

**Supp. Fig. S3**: Cardiometabolic analysis joint and interaction analysis results. The Manhattan plot displays association test strengths for the joint (top panel) and interaction (bottom panel) tests as a function of genomic position (*x*-axis).

****

**Supp. Fig. S4**: Multiple deprivation index analysis joint and interaction analysis results. The Manhattan plot displays association test strengths for the joint (top panel) and interaction (bottom panel) tests as a function of genomic position (*x*-axis).

****

**References**

Bach-Mortensen, A. M., & Degli Esposti, M. (2021). Is area deprivation associated with greater impacts of COVID-19 in care homes across England? A preliminary analysis of COVID-19 outbreaks and deaths. Journal of Epidemiology and Community Health, jech-2020-215039. doi:10.1136/jech-2020-215039

Bo Burström, Wenjing Tao, Social determinants of health and inequalities in COVID-19, European Journal of Public Health, Volume 30, Issue 4, August 2020, Pages 617–618, <https://doi.org/10.1093/eurpub/ckaa095>

Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes usingRegulomeDB. Genome Research 2012, 22(9):1790-1797. PMID: 22955989.

Ganna, A. (2021). Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. medRxiv, 2021.2003.2010.21252820. doi:10.1101/2021.03.10.21252820

Griffith, G. J., Morris, T. T., Tudball, M. J., Herbert, A., Mancano, G., Pike, L., . . . Hemani, G. (2020). Collider bias undermines our understanding of COVID-19 disease risk and severity. Nature Communications, 11(1), 5749. doi:10.1038/s41467-020-19478-2

Kamat, M. A. et al. (2019) PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics; 35(22):4851-4853

Klein, S., Flanagan, K. Sex differences in immune responses. Nat Rev Immunol 16, 626–638 (2016). <https://doi.org/10.1038/nri.2016.90>

Lau ES, McNeill JN, Paniagua SM, Liu EE, Wang JK, Bassett IV, et al. (2021) Sex differences in inflammatory markers in patients hospitalized with COVID-19 infection: Insights from the MGH COVID-19 patient registry. PLoS ONE 16(4): e0250774. <https://doi.org/10.1371/journal.pone.0250774>

Myers TA, Chanock SJ, Machiela MJ. (2020) LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. Front. Genet. 2020 Feb 28.

Ritter, A., Kreis, N.-N., Louwen, F., & Yuan, J. (2020). Obesity and COVID-19: Molecular Mechanisms Linking Both Pandemics. International Journal of Molecular Sciences, 21(16), 5793. Retrieved from https://www.mdpi.com/1422-0067/21/16/5793

Staley, J. R. et al. (2016) PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics; 32(20):3207-3209

Shastri A, Wheat J, Agrawal S, Chaterjee N, Pradhan K, Goldfinger M, et al. Delayed clearance of sars-cov2 in male compared to female patients: high ace2 expression in testes suggests possible existence of gender-specific viral reservoirs. MedRxiv. 2020; 2020.2004.2016.20060566.

The, C.-H. G. I. (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. European Journal of Human Genetics, 28(6), 715-718. doi:10.1038/s41431-020-0636-6

Udler MS, Kim J, von Grotthuss M, Bonàs-Guarch S, Cole JB, Chiou J, et al. (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med 15(9): e1002654. https://doi.org/10.1371/journal.pmed.1002654

Zhu, L., She, Z.-G., Cheng, X., Qin, J.-J., Zhang, X.-J., Cai, J., . . . Li, H. (2020). Association of Blood Glucose Control and Outcomes in Patients with COVID-19 and Pre-existing Type 2 Diabetes. Cell Metabolism, 31(6), 1068-1077.e1063. doi:https://doi.org/10.1016/j.cmet.2020.04.021