

Supplementary Material

Contents

1	Supplementary note	2
1.1	Genome-wide association study analyses (GWAS)	2
1.1.1	GenOMICC cohorts, multi-ancestry GWAS and meta-analysis	2
1.1.2	Multi-study GWAS meta-analysis	3
1.1.3	GenOMICC multi-ancestry fine-mapping and annotation	9
1.1.4	Age and Sex Stratified analysis	14
1.2	Polygenic risk score (PRS) analysis	15
1.3	Gene-based rare variant analysis tests (RVAT)	16
1.3.1	RVAT Cohorts	16
1.3.2	Multi-study RVAT meta-analysis	16
1.3.3	Individual variants contributing to the RVAT gene signals	19
1.3.4	Conditional analyses for <i>SLC50A1</i>	20
1.3.5	RVAT analysis stratified by age	20
2	Additional figures	23
3	Acknowledgments	32
4	Conflicts of Interest	33
5	Extended References	33

1 Supplementary note

This section presents additional analyses and detailed results that complement the main findings.

1.1 Genome-wide association study analyses (GWAS)

1.1.1 GenOMICC cohorts, multi-ancestry GWAS and meta-analysis

Following stringent quality control procedures (Methods), we conducted Genome-wide association analyses (GWAS) using SAIGE² for a cohort (GenOMICC) of 11,423 critically ill cases and 60,628 controls. 9,357 (82%) of the cases were part of the primary analysis in our prior publication¹. The controls included 100,000 Genomes Project (100kGP) individuals (n = 49,360) and recruited individuals with mild COVID-19 (n = 11,268) (Methods), mainly as part of the Real-time assessment of community transmission (REACT) study (cohort breakdown by ancestry shown in Extended Data Fig. 1).

GWAS analyses were carried out separately for four genetically inferred ancestry groups (African (AFR); East Asian (EAS); European (EUR); South Asian (SAS)) and using either all controls (ctrl-all) or only individuals with mild COVID-19 (ctrl-mld) with cohort breakdowns by ancestry shown in Extended Data Fig. 1. Demographic characteristics in terms of age and sex were broadly matched between mild COVID-19 and 100kGP control cohorts (Figure 1, Table 1). Severe cases and controls (ctrl-all) were broadly matched in terms of their ancestry composition (Figure 2). GWAS results were meta-analysed across ancestries and for each set of controls separately using inverse-variance-weighted fixed effects meta-analysis (IVW) (Methods).

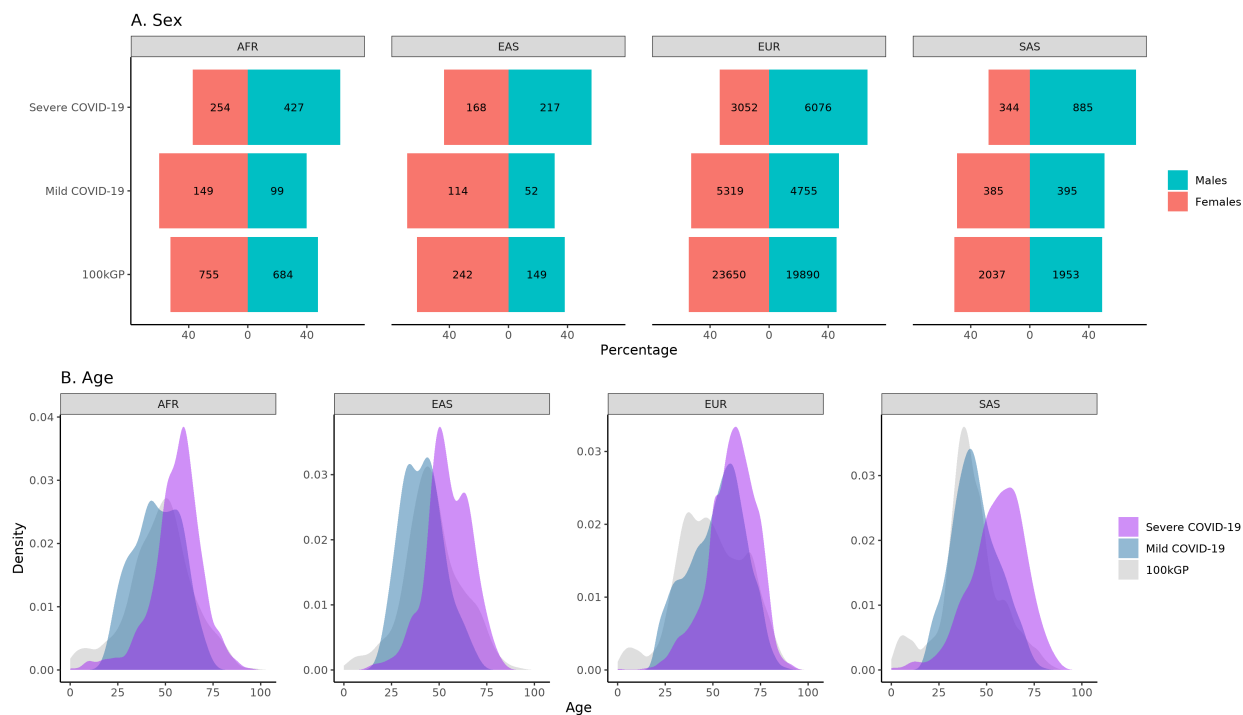


Figure 1: Demographic characteristics of analysed cohorts for the GenOMICC study

Cohort	All ancestries	AFR	EAS	EUR	SAS
All	51[26]	51[20]	49[19]	52[26]	43[20]
Severe COVID-19	60[17]	57.6[14]	54[15.8]	61[17]	57[19]
Mild COVID-19	53[22]	46[20]	41[16]	54[21.8]	43[17.2]
100kGP	48[26]	48[21]	45[18]	49[26]	41[15]

Table 1: Age medians and interquartile range in brackets across cohorts and ancestries.

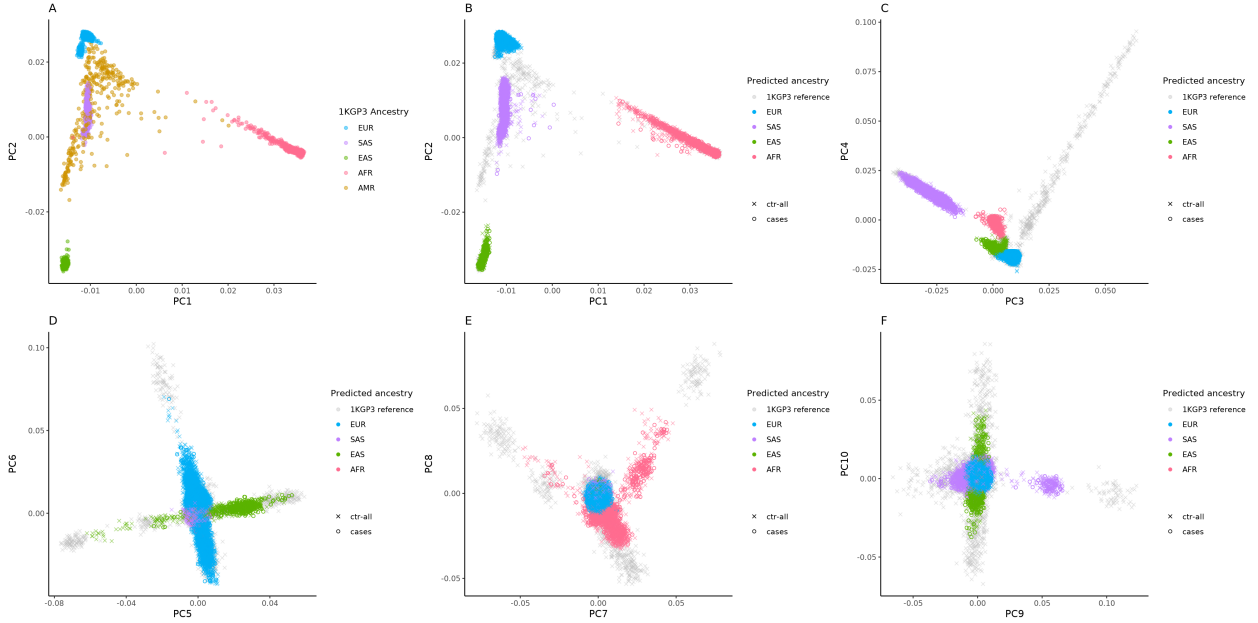


Figure 2: Projection of severe cases vs. ctrl-all samples onto 1,000 Genomes Project (1KGP3) principal components with coloring by predicted ancestry. Panel A shows the first two PCs for the 1KGP3 samples. Panels B-F show projections of the severe cases and ctrl-all samples of this study coloured by predicted ancestry, with PCs for 1KGP3 samples shown in grey.

1.1.2 Multi-study GWAS meta-analysis

To maximise power for GWAS discovery, we meta-analysed the GWAS summary statistics from three studies: GenOMICC severe versus ctrl-all (this study), HGIv7-A2_ALL.leave.23andme.and.genomicc², and 23andMe EUR respiratory support³ (included cohorts breakdown in Extended Data Fig. 1). To account for potential heterogeneity in variant overlap and ancestry composition across studies that could confound conditional analysis on the basis of meta-analysis summaries⁴, we defined locus associations as $\pm 1\text{Mbp}$ regions surrounding top variant signals (Methods). We identified 38 genome-wide significant loci ($P < 5 \times 10^{-8}$), of which five were not previously reported, near *ARF1*, *MECOM*, *ZNF462*, *KLF13* and *MVP* genes (summary statistics in Table 2, allele frequencies in Table ??)

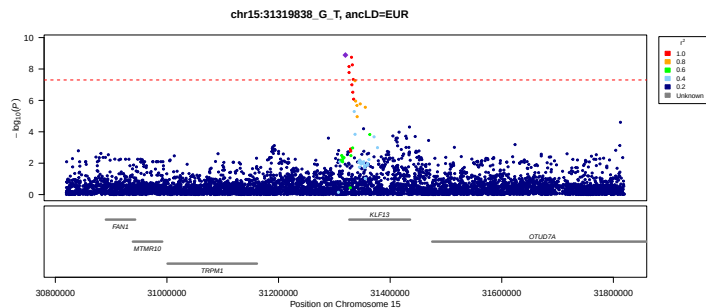
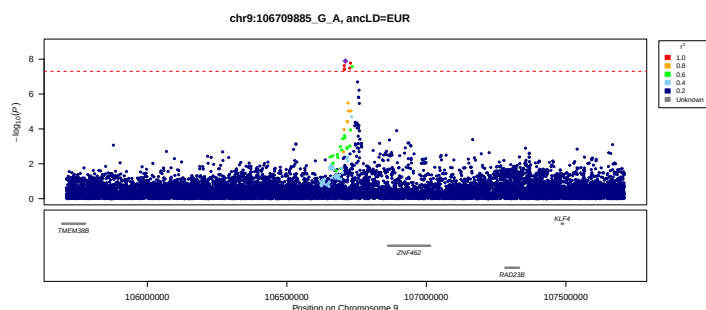
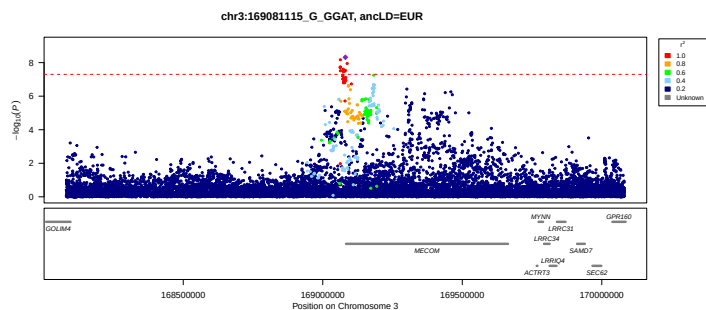
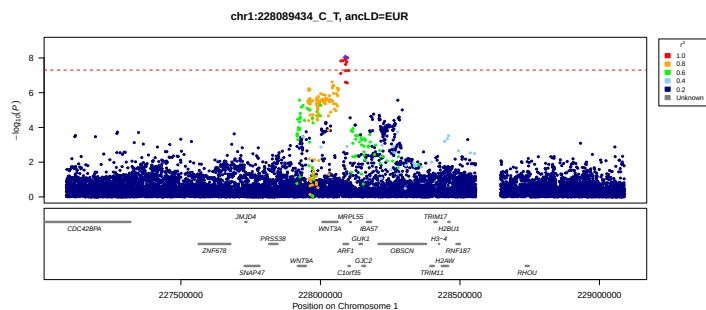
Each of the novel lead variant signals for *ARF1*, *MECOM*, *ZNF462*, *KLF13* was supported by multiple genome-wide significant variants ($P < 5 \times 10^{-8}$) in strong LD with the lead (EUR $r_{EUR}^2 > 0.8$; Figure 3). *MVP* was supported by three variants with $P < 10^{-4}$ in moderate LD with the lead ($r_{AFR}^2 > 0.4$; Figure 3). Additionally, lead variant odds ratios for all signals were consistent with meta-analysis using only the ctrl-mld subset of controls (Table 4, Figure 4, and Figure 3).

We further assessed between-study heterogeneity and support from the second largest study (HGI) to increase confidence in these associations. For the latter, we considered a $P < 0.0125$ (taken as $P = 0.05$ corrected for the number of assessed signals) with the same direction of effect in the HGI cohort as evidence supporting the respective signal. Signals near *ARF1*, *ZNF462*, *KLF13* had support across studies and no evidence of heterogeneity, but *MECOM* has $P > 0.0125$ in the HGI cohort and $P_{het} = 2.13 \times 10^{-3}$ (Table 1, main manuscript). We therefore consider the signal near *MECOM* as requiring further replication in an independent cohort.

The locus near *MVP* could not be assessed because no other studies had sufficient data. The lead variant is very rare in non-African ancestries (maf of 1.3×10^{-4} in gnomAD NFE, v4.1).

Chr:pos(b38):ref:alt	rsid	OR	OR_{CI}	P_{meta}	P_{het}	$P_{GenOMICC}$	P_{hgi}	Nearest gene	Citation
1:9062610:A:G	rs2505973	0.932	0.91-0.954	4.00×10^{-9}	4.32×10^{-1}	3.82×10^{-7}	1.63×10^{-3}	<i>SLC2A5</i>	GenOMICC3
1:64951326:A:G*	rs12730021	0.897	0.871-0.923	3.90×10^{-13}	3.52×10^{-1}	2.54×10^{-10}	1.54×10^{-4}	<i>JAK1</i>	GenOMICC3
1:155197995:A:G*	rs41264915	0.823	0.793-0.854	8.62×10^{-25}	5.08×10^{-1}	1.91×10^{-16}	9.71×10^{-10}	<i>THBS3</i>	GenOMICC2
1:228089434:C:T	rs3738681	0.881	0.843-0.92	9.30×10^{-9}	2.55×10^{-1}	1.00×10^{-5}	1.10×10^{-3}	<i>ARF1</i>	Novel
2:60480453:A:G*	rs1123573	0.904	0.882-0.926	9.77×10^{-16}	3.22×10^{-1}	1.41×10^{-9}	2.81×10^{-8}	<i>BCL11A</i>	GenOMICC2
3:45818159:G:A*	rs17713054	2.05	1.97-2.14	8.34×10^{-266}	3.49×10^{-10}	2.18×10^{-192}	1.14×10^{-77}	<i>LZTFL1</i>	SCGG19
3:146517122:G:A*	rs343320	1.15	1.1-1.21	1.30×10^{-9}	7.23×10^{-2}	1.72×10^{-9}	3.38×10^{-2}	<i>PLSCR1</i>	GenOMICC2
3:169081115:G:GGAT*	rs112039646	1.15	1.1-1.2	4.81×10^{-9}	-	4.58×10^{-9}	-	<i>MECOM</i>	Novel
4:25446871:A:G	rs7664615	0.906	0.879-0.933	3.27×10^{-11}	2.59×10^{-1}	2.39×10^{-7}	3.11×10^{-4}	<i>ANAPC4</i>	GenOMICC3
4:105897896:G:A*	rs34712979	0.902	0.875-0.929	9.31×10^{-12}	2.12×10^{-1}	2.23×10^{-9}	4.15×10^{-4}	<i>NPNT</i>	GenOMICC3
5:132422622:A:G*	rs2269821	0.908	0.879-0.938	5.21×10^{-9}	2.27×10^{-1}	1.56×10^{-7}	1.86×10^{-3}	<i>IRF1</i>	GenOMICC2
6:31182658:A:G	rs1128175	0.889	0.865-0.914	1.11×10^{-16}	2.15×10^{-1}	3.80×10^{-8}	1.98×10^{-8}	<i>HLA</i>	GenOMICC1
6:41520640:G:A*	rs12660421	1.41	1.33-1.5	7.99×10^{-29}	2.24×10^{-1}	2.26×10^{-18}	2.38×10^{-11}	<i>FOXP4</i>	HGI21
7:75634474:C:T	rs12534422	1.08	1.05-1.11	9.90×10^{-10}	6.37×10^{-1}	5.56×10^{-7}	5.78×10^{-4}	<i>HIP1</i>	GenOMICC3
7:100032719:C:T*	rs2897075	1.09	1.06-1.11	7.31×10^{-13}	6.55×10^{-1}	2.15×10^{-8}	1.35×10^{-5}	<i>ZKSCAN1</i>	GenOMICC3
8:60568368:GTC:G	rs34038069	1.07	1.05-1.1	7.00×10^{-9}	5.10×10^{-1}	1.84×10^{-6}	7.13×10^{-4}	<i>RAB2A</i>	GenOMICC3
9:21206606:C:G*	rs28368148	1.54	1.38-1.73	7.95×10^{-14}	2.76×10^{-1}	4.51×10^{-11}	1.10×10^{-4}	<i>IFNA10</i>	GenOMICC2
9:106709885:G:A	rs60568503	1.07	1.05-1.1	1.28×10^{-8}	6.97×10^{-1}	4.65×10^{-6}	4.61×10^{-4}	<i>ZNF462</i>	Novel
9:133271182:T:C*	rs550057	0.889	0.866-0.912	1.04×10^{-18}	7.73×10^{-1}	7.57×10^{-9}	1.24×10^{-10}	<i>ABO</i>	SCGG19
10:79946568:A:G*	rs721917	1.09	1.06-1.11	1.87×10^{-13}	1.01×10^{-1}	4.00×10^{-8}	1.77×10^{-5}	<i>SFTPD</i>	HGI22
11:1214934:A:G	rs12802931	0.899	0.87-0.929	1.54×10^{-10}	7.17×10^{-1}	1.98×10^{-6}	1.15×10^{-5}	<i>MUC5B</i>	HGI22
11:34482745:G:A*	rs61882275	0.885	0.864-0.906	2.29×10^{-24}	2.44×10^{-1}	3.42×10^{-17}	2.85×10^{-9}	<i>ELF5</i>	GenOMICC2
12:112919388:G:A*	rs10774671	1.11	1.08-1.14	9.89×10^{-16}	9.61×10^{-1}	1.06×10^{-9}	1.35×10^{-7}	<i>OAS1</i>	GenOMICC1
12:132481571:G:A*	rs11614702	1.11	1.08-1.13	3.71×10^{-18}	6.04×10^{-2}	1.27×10^{-14}	2.36×10^{-6}	<i>FBRSL1</i>	GenOMICC2
13:112881427:C:T*	rs12585036	1.14	1.11-1.17	1.14×10^{-20}	9.50×10^{-2}	3.87×10^{-16}	6.72×10^{-7}	<i>ATP11A</i>	GenOMICC2
15:31319838:G:T*	rs11636034	0.924	0.9-0.948	1.30×10^{-9}	1.24×10^{-1}	3.68×10^{-8}	2.47×10^{-3}	<i>KLF13</i>	Novel
16:29843264:G:A*	rs138640006	2.22	1.67-2.94	2.66×10^{-8}	-	2.53×10^{-8}	-	<i>MVP</i>	Novel
16:89196249:G:A*	rs117169628	1.16	1.12-1.2	3.07×10^{-17}	2.70×10^{-1}	2.44×10^{-9}	1.26×10^{-7}	<i>SLC22A31</i>	GenOMICC2
17:40003082:T:C*	rs12941811	1.08	1.05-1.1	3.82×10^{-10}	3.50×10^{-2}	1.89×10^{-8}	7.05×10^{-3}	<i>PSMD3</i>	GenOMICC3
17:46779515:G:A	rs199514	1.11	1.08-1.15	9.70×10^{-13}	5.95×10^{-1}	5.35×10^{-6}	2.31×10^{-8}	<i>WNT3</i>	Degenhardt
17:49863260:C:A*	rs3848456	1.32	1.24-1.4	8.79×10^{-20}	3.92×10^{-1}	8.95×10^{-12}	4.37×10^{-10}	<i>TAC4</i>	Degenhardt
19:4717660:A:G*	rs12610495	1.26	1.23-1.29	6.13×10^{-67}	2.59×10^{-2}	1.70×10^{-48}	6.52×10^{-22}	<i>DPP9</i>	GenOMICC1
19:10352442:G:C*	rs34536443	1.48	1.39-1.58	4.69×10^{-35}	5.93×10^{-1}	6.23×10^{-25}	2.05×10^{-10}	<i>TYK2</i>	GenOMICC1
19:48702888:G:C*	rs516316	0.914	0.893-0.936	2.50×10^{-14}	8.50×10^{-2}	3.69×10^{-12}	1.28×10^{-4}	<i>FUT2</i>	GenOMICC2
20:6489447:G:A	rs2326788	0.928	0.905-0.951	2.14×10^{-9}	9.06×10^{-1}	7.04×10^{-6}	1.89×10^{-4}	<i>BMP2</i>	GenOMICC3
21:33240996:A:G*	rs17860169	1.2	1.17-1.23	1.83×10^{-49}	1.80×10^{-1}	2.15×10^{-32}	6.18×10^{-19}	<i>IFNAR2</i>	GenOMICC1
21:41471061:G:A*	rs9305744	0.909	0.885-0.934	5.48×10^{-12}	7.83×10^{-1}	4.13×10^{-8}	3.85×10^{-5}	<i>TMPPRSS2</i>	GenOMICC3
X:15602217:T:C*	rs190509934	0.465	0.373-0.58	1.02×10^{-11}	4.45×10^{-1}	2.92×10^{-9}	5.26×10^{-4}	<i>ACE2</i>	HGI22

Table 2: Lead variants of locus association signals in the multi-study GWAS meta-analysis. Variant ids are provided as Chr:pos(b38):ref:alt using GRCh38 coordinates with ref indicating the reference allele and alt being the effect allele and an asterisk (*) indicates that the locus indexed by the lead variant is fine-mapped. OR , OR_{CI} and P_{meta} are Odds ratios, Odds ratio 95% confidence intervals, and the P -value from the meta-analysis. P_{het} is the P -value from heterogeneity analysis. $P_{GenOMICC}$ and P_{hgi} are P -values for lead variants from the GenOMICC multi-ancestry meta-analysis and HGIv7-leaveGenOMICC data, respectively. Citation column refers to the first study reporting a significant signal within ± 1 Mbp of the lead variant of this study, with “Novel” indicating novel significant loci identified in this study (Referenced studies being either GenOMICC1¹, GenOMICC2⁵, GenOMICC3¹, SCGG19⁶, HGI21⁷, HGI22⁸, Degenhardt⁹). Signals within the extended MHC region are annotated as “HLA”.



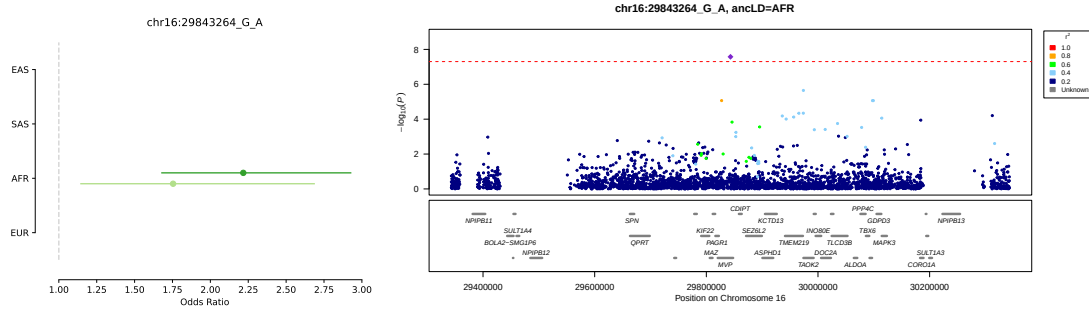


Figure 3: LocusZoom and forest plots illustrating the novel genetic association signals identified in this study. Panels on the right compare cross-ancestry odds ratios of lead variants from analyses using all controls (ctrl-all, darker lines) vs. analyses using only controls with mild COVID-19 (ctrl-mld, lighter lines). Panels on the right show LocusZoom plots centered on lead variant. P -values are from the cross-ancestry GenOMICC severe versus ctrl-all meta-analysis. The population used to calculate LD is indicated on the title and corresponds to the population having the smallest P -value in the GenOMICC severe vs ctrl-all meta-analysis. Equivalent plots for all lead variants in Table 2 are in Figure 9.

Chr:pos(b38):ref:alt	rsid	OR	OR_{CI}	P	$OR_{ctrl-mld}$	$OR_{CI_{ctrl-mld}}$	$P_{ctrl-mld}$
1:9062610:A:G	rs2505973	0.932	0.91-0.954	4.00×10^{-9}	0.906	0.867-0.946	8.99×10^{-6}
1:64951326:A:G	rs12730021	0.897	0.871-0.923	3.90×10^{-13}	0.931	0.884-0.981	7.49×10^{-3}
1:155197995:A:G	rs41264915	0.823	0.793-0.854	8.62×10^{-25}	0.757	0.707-0.811	2.66×10^{-15}
1:228089434:C:T	rs3738681	0.881	0.843-0.92	9.30×10^{-9}	0.861	0.794-0.934	2.84×10^{-4}
2:60480453:A:G	rs1123573	0.904	0.882-0.926	9.77×10^{-16}	0.919	0.879-0.96	1.66×10^{-4}
3:45818159:G:A	rs17713054	2.05	1.97-2.14	8.34×10^{-266}	2.06	1.93-2.2	2.36×10^{-100}
3:146517122:G:A	rs343320	1.15	1.1-1.21	1.30×10^{-9}	1.17	1.08-1.26	7.58×10^{-5}
3:169081115:G:GGAT	rs112039646	1.15	1.1-1.2	4.81×10^{-9}	1.13	1.06-1.2	1.17×10^{-4}
4:25446871:A:G	rs7664615	0.906	0.879-0.933	3.27×10^{-11}	0.897	0.851-0.945	4.44×10^{-5}
4:105897896:G:A	rs34712979	0.902	0.875-0.929	9.31×10^{-12}	0.916	0.87-0.964	7.60×10^{-4}
5:132422622:A:G	rs2269821	0.908	0.879-0.938	5.21×10^{-9}	0.882	0.829-0.94	9.27×10^{-5}
6:31182658:A:G	rs1128175	0.889	0.865-0.914	1.11×10^{-16}	0.916	0.871-0.963	5.40×10^{-4}
6:41520640:G:A	rs12660421	1.41	1.33-1.5	7.99×10^{-29}	1.43	1.27-1.61	4.88×10^{-9}
7:75634474:C:T	rs12534422	1.08	1.05-1.11	9.90×10^{-10}	1.1	1.05-1.15	5.44×10^{-5}
7:100032719:C:T	rs2897075	1.09	1.06-1.11	7.31×10^{-13}	1.08	1.03-1.13	7.37×10^{-4}
8:60568368:GTC:G	rs34038069	1.07	1.05-1.1	7.00×10^{-9}	1.04	1-1.09	4.40×10^{-2}
9:21206606:C:G	rs28368148	1.54	1.38-1.73	7.95×10^{-14}	1.58	1.32-1.89	7.35×10^{-7}
9:106709885:G:A	rs60568503	1.07	1.05-1.1	1.28×10^{-8}	1.09	1.04-1.14	9.34×10^{-5}
9:133271182:T:C	rs550057	0.889	0.866-0.912	1.04×10^{-18}	0.936	0.892-0.982	7.47×10^{-3}
10:79946568:A:G	rs721917	1.09	1.06-1.11	1.87×10^{-13}	1.08	1.04-1.13	2.03×10^{-4}
11:1214934:A:G	rs12802931	0.899	0.87-0.929	1.54×10^{-10}	0.904	0.855-0.956	3.98×10^{-4}
11:34482745:G:A	rs61882275	0.885	0.864-0.906	2.29×10^{-24}	0.88	0.842-0.919	9.98×10^{-9}
12:112919388:G:A	rs10774671	1.11	1.08-1.14	9.89×10^{-16}	1.12	1.07-1.17	1.18×10^{-6}
12:132481571:G:A	rs11614702	1.11	1.08-1.13	3.71×10^{-18}	1.13	1.08-1.18	2.03×10^{-8}
13:112881427:C:T	rs12585036	1.14	1.11-1.17	1.14×10^{-20}	1.17	1.11-1.23	2.84×10^{-9}
15:31319838:G:T	rs11636034	0.924	0.9-0.948	1.30×10^{-9}	0.921	0.879-0.964	4.58×10^{-4}
16:29843264:G:A	rs138640006	2.22	1.67-2.94	2.66×10^{-8}	1.75	1.14-2.71	1.11×10^{-2}
16:89196249:G:A	rs117169628	1.16	1.12-1.2	3.07×10^{-17}	1.12	1.06-1.19	1.34×10^{-4}
17:40003082:T:C	rs12941811	1.08	1.05-1.1	3.82×10^{-10}	1.06	1.01-1.1	1.57×10^{-2}
17:46779515:G:A	rs199514	1.11	1.08-1.15	9.70×10^{-13}	1.06	1-1.12	4.53×10^{-2}
17:49863260:C:A	rs3848456	1.32	1.24-1.4	8.79×10^{-20}	1.22	1.1-1.36	2.24×10^{-4}
19:4717660:A:G	rs12610495	1.26	1.23-1.29	6.13×10^{-67}	1.27	1.21-1.33	1.01×10^{-25}
19:10352442:G:C	rs34536443	1.48	1.39-1.58	4.69×10^{-35}	1.48	1.34-1.62	4.23×10^{-16}
19:48702888:G:C	rs516316	0.914	0.893-0.936	2.50×10^{-14}	0.898	0.861-0.938	8.94×10^{-7}
20:6489447:G:A	rs2326788	0.928	0.905-0.951	2.14×10^{-9}	0.948	0.907-0.99	1.62×10^{-2}
21:33240996:A:G	rs17860169	1.2	1.17-1.23	1.83×10^{-49}	1.23	1.17-1.28	1.70×10^{-19}
21:41471061:G:A	rs9305744	0.909	0.885-0.934	5.48×10^{-12}	0.892	0.848-0.937	6.94×10^{-6}
X:15602217:T:C	rs190509934	0.465	0.373-0.58	1.02×10^{-11}	0.239	0.105-0.543	6.36×10^{-4}

Table 4: Comparison of effect sizes of lead variants from the multi-study meta-analysis versus a multi-ancestry meta-analysis of GenOMICC cases using only individuals with mild COVID-19 as controls (GenOMICC severe versus ctrl-mld). Variant ids are provided as Chr:pos(b38):ref:alt using GRCh38 coordinates with ref indicating the reference allele and alt being the effect allele. OR , OR_{CI} and P are Odds ratios, Odds ratio 95% confidence intervals and P -values for the multi-study meta analysis. OR , OR_{CI} and P with subscript $ctrl - mld$ are derived from the meta-analysis of the GenOMICC severe vs. ctrl-mld GWAS.

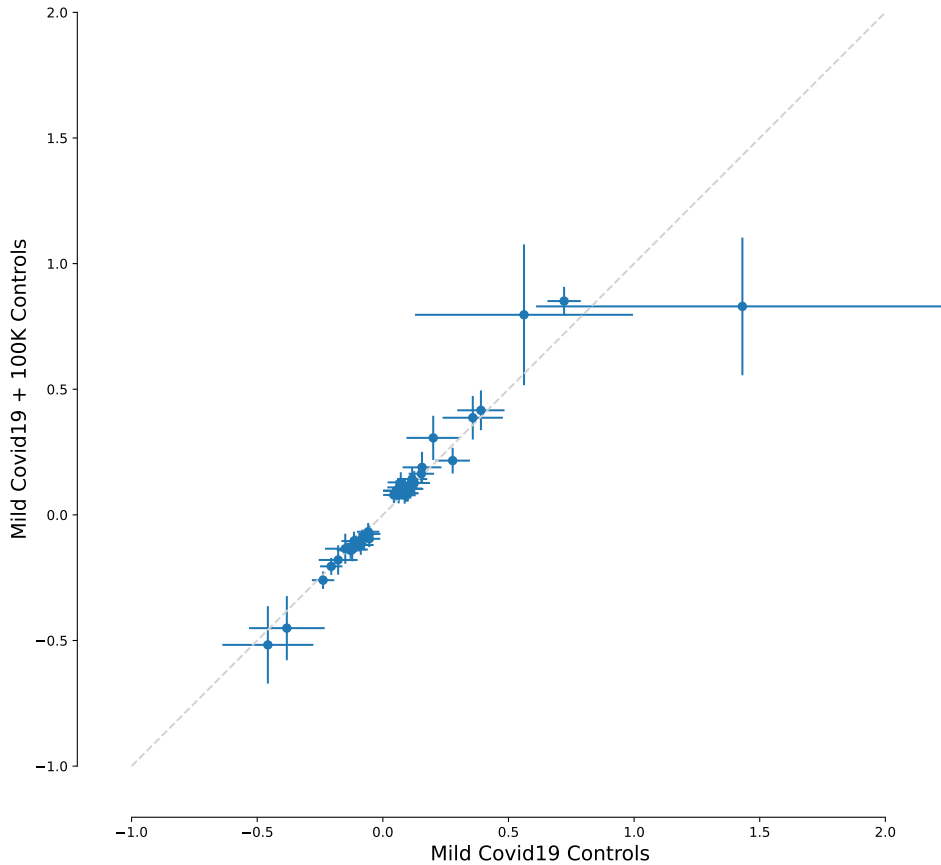


Figure 4: Comparison of logOR estimates at lead variants using either only mild Covid19 cases as controls (GenOMICC severe versus ctrl-mld) or, as in the main analysis, both mild Covid19 cases and 100K Genomes participants as controls (GenOMICC severe versus ctrl-all).

1.1.3 GenOMICC multi-ancestry fine-mapping and annotation

We performed multi-ancestry fine-mapping to identify credible sets (CS) with MESuSiE¹⁰ for a subset of 28 loci that contained significant variant signals ($P < 5 \times 10^{-8}$) in the GenOMICC severe versus ctrl-all analysis (loci indicated with asterisk on variant ID index; Table 2). We fine-mapped a total of 64 CS that were distributed within the 28 locus signals (Table 5). We annotated each CS lead variant with the best scoring V2G gene from the OpenTargets Genetics Platform and VEP predicted consequence for the canonical transcript (Table 5). Among the list of predicted consequences, we identified missense variants for *PLSCR1*, *IFNA10*, *SLC22A31* and *TYK2* genes (Table 5). We annotated additional consequences for CS variants by identifying the worst predicted VEP consequence across the CS to find an additional two missense variants at *THBS3* and *GSDMA* genes (Table 6). We also performed fine-mapping for the EUR ancestry GenOMICC cohort using SUSIE¹¹ which yielded a total of 58 credible sets that largely matched the multi-ancestry fine-mapping results in terms of prioritised genes and consequences (Table 7). EUR fine-mapping identified an

additional missense variant for *IFNAR2* (Table 8).

Locus index variant	iCS	nCS	CS lead variant	PIP	MinP	MinP pop	V2G gene	CADD	VEP consq	VEP gene
1:64951326:A:G	1	1	1:64951326:A:G	1	8.38×10^{-11}	META	<i>JAK1</i>	2.3	intron	<i>JAK1</i>
1:64951326:A:G	2	2	1:64964168:T:C	0.57	2.68×10^{-7}	META	<i>JAK1</i>	7	intron	<i>JAK1</i>
1:155197995:A:G	1	1	1:155066988:C:T	0.95	1.28×10^{-18}	META	<i>EFNA4</i>	10	synonymous	<i>EFNA4</i>
1:155197995:A:G	2	3	1:155197995:A:G	0.5	2.42×10^{-17}	META	<i>MUC1</i>	2.1	intron	<i>THBS3</i>
1:155197995:A:G	3	20	1:155278322:A:T	0.68	9.60×10^{-6}	META	<i>FDP5</i>	5.2	intron	<i>HCN3</i>
1:155197995:A:G	4	4	1:155175305:G:A	0.48	7.46×10^{-10}	META	<i>MUC1</i>	8.3	upstream	<i>KRTCAP2</i>
2:60480453:A:G	1	7	2:60480759:G:A	0.71	2.75×10^{-9}	META	<i>BCL11A</i>	7.4	intron	<i>BCL11A</i>
3:45818159:G:A	1	5	3:45859597:C:T	0.83	4.70×10^{-204}	META	<i>CXCR6</i>	0.14	intron	-
3:45818159:G:A	2	1	3:45329214:G:A	1	1.36×10^{-62}	EUR	<i>LARS2</i>	0.34	-	-
3:45818159:G:A	3	1	3:45796521:G:T	1	4.69×10^{-18}	EUR	<i>FYCO1</i>	9.2	5'UTR	<i>SLC6A20</i>
3:45818159:G:A	4	1	3:45239823:G:A	1	4.07×10^{-1}	SAS	<i>TMEM158</i>	0.68	-	-
3:45818159:G:A	5	13	3:46360250:G:A	0.1	1.35×10^{-1}	AFR	<i>CCR1</i>	4.7	3'UTR	<i>CCR2</i>
3:45818159:G:A	6	1	3:45280485:C:T	1	5.03×10^{-1}	EUR	<i>TMEM158</i>	7.6	-	-
3:45818159:G:A	7	2	3:45645361:AAC:A	0.66	8.56×10^{-2}	SAS	<i>SACM1L</i>	0.66	intron	<i>LIMD1</i>
3:45818159:G:A	8	2	3:45570834:G:C	0.5	4.27×10^{-1}	SAS	<i>LIMD1</i>	11	-	-
3:45818159:G:A	9	2	3:45592155:A:T	0.52	2.27×10^{-1}	AFR	<i>LIMD1</i>	1.5	upstream	<i>LIMD1</i>
3:146517122:G:A	1	9	3:146517122:G:A	0.17	2.28×10^{-10}	EUR	<i>PLSCR1</i>	23	missense	<i>PLSCR1</i>
3:169081115:G:GGAT	1	36	3:169081067:C:T	0.061	1.44×10^{-9}	META	<i>MECOM</i>	0.35	downstream	<i>MECOM</i>
3:169081115:G:GGAT	2	9	3:169521790:A:G	0.16	1.28×10^{-6}	EUR	<i>MECOM</i>	1.9	intron	<i>MECOM</i>
3:169081115:G:GGAT	3	6	3:169299683:T:C	0.54	1.30×10^{-6}	EUR	<i>MECOM</i>	4	intron	<i>MECOM</i>
3:169081115:G:GGAT	4	1	3:169515899:A:AAATT	1	2.32×10^{-7}	META	<i>MECOM</i>	0.67	intron	<i>MECOM</i>
4:105897896:G:A	1	1	4:105897896:G:A	1	4.65×10^{-10}	EUR	<i>NPNT</i>	24	spliceacceptor	<i>NPNT</i>
4:105897896:G:A	2	1	4:105554756:C:CAT	1	5.78×10^{-8}	AFR	<i>PPA2</i>	0.74	intron	<i>ARHGEF38</i>
5:132422622:A:G	1	22	5:131995059:C:T	0.47	7.23×10^{-10}	EUR	<i>ACSL6</i>	0.21	intron	<i>ACSL6</i>
7:100032719:C:T	1	4	7:100032719:C:T	0.5	8.43×10^{-9}	META	<i>TRIM4</i>	1.8	intron	<i>ZKSCAN1</i>
9:21206606:C:G	1	3	9:21206606:C:G	0.73	1.11×10^{-11}	EUR	<i>IFNA10</i>	24	missense	<i>IFNA10</i>
9:21206606:C:G	2	2	9:21328701:G:A	0.53	8.81×10^{-9}	META	<i>KLHL9</i>	11	downstream	<i>KLHL9</i>
9:133271182:T:C	1	9	9:133271182:T:C	0.84	2.82×10^{-9}	META	-	4.4	intron	<i>ABO</i>
11:34482745:G:A	1	3	11:34482745:G:A	0.58	5.13×10^{-18}	META	<i>CAT</i>	0.073	intron	<i>ELF5</i>
12:112919388:G:A	1	60	12:112919388:G:A	0.15	3.79×10^{-10}	META	<i>OAS1</i>	4.2	spliceacceptor	<i>OAS1</i>
12:132481571:G:A	1	13	12:132479849:C:T	0.14	1.34×10^{-13}	META	<i>FBRSL1</i>	0.94	-	-
12:132481571:G:A	2	1	12:132158547:G:A	1	3.86×10^{-12}	EUR	<i>DDX51</i>	0.31	-	-
13:112881427:C:T	1	3	13:112886111:C:T	0.4	3.31×10^{-17}	META	<i>ATP11A</i>	5.5	3'UTR	<i>ATP11A</i>
15:31319838:G:T	1	1	15:31334071:C:T	0.99	1.34×10^{-8}	EUR	<i>KLF13</i>	5.6	intron	<i>KLF13</i>
15:31319838:G:T	2	9	15:31319838:G:T	0.22	1.48×10^{-8}	META	<i>KLF13</i>	2.7	-	-
16:29843264:G:A	1	2	16:29843264:G:A	0.93	1.84×10^{-8}	AFR	<i>MVP</i>	0.46	intron	<i>MVP</i>
16:29843264:G:A	2	12	16:29709545:G:A	0.11	4.16×10^{-6}	EUR	<i>MVP</i>	1.1	downstream	-
16:89196249:G:A	1	2	16:89196249:G:A	0.72	4.67×10^{-10}	EUR	<i>SLC22A31</i>	23	missense	<i>SLC22A31</i>
17:40003082:T:C	1	66	17:40003082:T:C	0.72	1.41×10^{-9}	META	<i>MED24</i>	2.2	-	-
17:49863260:C:A	1	3	17:49863260:C:A	0.75	9.48×10^{-13}	META	<i>DLX3</i>	5.4	-	-
17:49863260:C:A	2	1	17:50213197:T:C	1	1.41×10^{-9}	META	<i>TMEM92</i>	2.7	intron	-
17:49863260:C:A	3	2	17:50212932:G:A	0.88	4.33×10^{-1}	SAS	<i>TMEM92</i>	1.6	intron	-
17:49863260:C:A	4	1	17:50030360:G:A	0.96	1.58×10^{-7}	EUR	<i>ITGA3</i>	0.29	-	-
17:49863260:C:A	5	2	17:50219800:G:A	0.19	4.71×10^{-1}	SAS	<i>TMEM92</i>	12	downstream	-
17:49863260:C:A	6	1	17:50109066:C:T	1	3.11×10^{-1}	EAS	<i>ITGA3</i>	2.5	downstream	<i>SAMD14</i>
19:4717660:A:G	1	1	19:4717660:A:G	1	3.04×10^{-51}	META	<i>DPP9</i>	16	intron	<i>DPP9</i>
19:4717660:A:G	2	1	19:5066102:C:T	1	1.87×10^{-14}	EUR	<i>KDM4B</i>	0.2	intron	<i>KDM4B</i>
19:10352442:G:C	1	1	19:10352442:G:C	0.97	1.75×10^{-26}	META	<i>TYK2</i>	25	missense	<i>TYK2</i>
19:10352442:G:C	2	1	19:10414696:G:A	1	1.70×10^{-11}	META	<i>ICAM1</i>	1.1	-	-
19:48702888:G:C	1	1	19:48496300:TC:T	1	5.66×10^{-15}	META	<i>LMTK3</i>	0.78	intron	<i>LMTK3</i>
19:48702888:G:C	2	10	19:48697960:C:T	0.42	5.18×10^{-13}	META	<i>RASIP1</i>	2.4	intron	<i>FUT2</i>
19:48702888:G:C	3	1	19:48500758:G:A	1	1.14×10^{-1}	META	<i>LMTK3</i>	8	intron	<i>LMTK3</i>
19:48702888:G:C	4	2	19:48494890:G:A	0.48	3.32×10^{-2}	META	<i>LMTK3</i>	0.78	intron	<i>LMTK3</i>
21:33240996:A:G	1	1	21:33412149:C:T	1	1.53×10^{-37}	EUR	<i>IFNAR2</i>	2.2	intron	<i>IFNGR2</i>
21:33240996:A:G	2	1	21:34355721:GGAAA:G	1	1.29×10^{-33}	EUR	<i>ATP5PO</i>	0.63	intron	-
21:33240996:A:G	3	9	21:33262748:G:GT	0.73	8.91×10^{-36}	META	<i>IFNAR2</i>	0.85	upstream	<i>IL10RB</i>
21:33240996:A:G	4	11	21:33954962:A:G	0.19	4.58×10^{-18}	META	<i>MRPS6</i>	3.8	intron	<i>LINC00649</i>
21:33240996:A:G	5	1	21:33441927:G:A	1	1.97×10^{-2}	EUR	<i>IFNAR2</i>	0.48	downstream	<i>IFNGR2</i>
21:33240996:A:G	6	30	21:33287696:C:T	0.1	2.32×10^{-7}	EUR	<i>IL10RB</i>	1.4	intron	<i>IL10RB</i>
21:41471061:G:A	1	24	21:41481267:TCAGACA:T	0.28	4.07×10^{-8}	META	<i>MX1</i>	0.73	intron	<i>TMPRSS2</i>
X:15602217:T:C	1	1	X:15752522:A:G	1	6.95×10^{-12}	EUR	<i>CA5B</i>	2.6	intron	<i>CA5B</i>
X:15602217:T:C	2	1	X:15602217:T:C	1	5.10×10^{-14}	META	<i>ACE2</i>	8.6	upstream	<i>ACE2</i>
X:15602217:T:C	3	1	X:15644354:C:T	1	5.08×10^{-9}	EUR	<i>CLTRN</i>	2.5	downstream	-
X:15602217:T:C	4	14	X:15616792:T:G	0.12	9.91×10^{-8}	AFR	<i>ACE2</i>	3	intron	<i>GSI-594A7.3</i>

Table 5: Lead variants in credible sets from **multi-ancestry** fine-mapping in GenOMICC severe versus ctrl-all cohort. Locus index variant indicates index variant for locus as identified by the multi-study meta-analysis and is the variant with the highest posterior inclusion probability (PIP) in each credible set. *iCS* and *nCS* indicate the index and the number of variants in each credible set. *PIP* is the posterior inclusion probability of the lead variant. *MinP* indicates the minimum *P*-value for the lead variant among AFR, EAS, EUR, SAS or META from GenOMICC per population GWASs and multi-ancestry meta-analysis, while *MinP pop* indicates the analysis that produced the lowest *P*-value. *V2G gene* is the gene prioritised with the highest score by the Variant-to-Gene (V2G) pipeline on the Open Targets Genetics platform for the lead variant. *CADD* is the Combined Annotation Dependent Depletion score for the lead variant while *VEP consq* and *VEP gene* indicate the VEP predicted consequence on the canonical transcript and impacted gene, respectively.

Locus index variant	iCS	nCS	CS worst variant	PIP	MinP	MinP pop	CADD	VEP consq	VEP gene
1:155197995:A:G	2	3	1:155202934:T:C	0.33	6.10×10^{-17}	META	21	missense	<i>THBS3</i>
1:155197995:A:G	3	20	1:155260340:C:T	0.005	2.85×10^{-4}	META	2.5	synonymous	<i>SCAMP3</i>
17:40003082:T:C	1	66	17:39974934:C:A	0.0023	1.27×10^{-6}	EUR	22	missense	<i>GSDMA</i>
19:48702888:G:C	2	10	19:48703346:C:T	0.1	1.43×10^{-12}	META	7	synonymous	<i>FUT2</i>
21:41471061:G:A	1	24	21:41473456:A:G	0.0036	1.42×10^{-6}	META	1.8	synonymous	<i>TMPRSS2</i>

Table 6: Variants with worst consequence in credible sets from **multi-ancestry** fine-mapping in the GenOMICC severe versus ctrl-all cohort. Variants with the worst VEP consequence in each fine-mapped credible are shown, for cases when the worst variant is not the lead variant and the consequence has low impact or worse (“LOW”, “MODERATE” or “HIGH” in VEP impact scale). *PIP*, *MinP*, *MinP pop*, *CADD*, *VEP consq* and *VEP gene* columns refer to the CS worst variant. *MinP* indicates the minimum *P*-value among AFR, EAS, EUR, SAS or META from GenOMICC per population GWASs and multi-ancestry meta-analysis, while *MinP pop* indicates the analysis that produced the lowest *P*-value. *CADD* is the Combined Annotation Dependent Depletion score while *VEP consq* and *VEP gene* indicate the VEP predicted consequence on the canonical transcript and impacted gene, respectively.

Locus index variant	iCS	nCS	CS lead variant	PIP	P	V2G gene	CADD	VEP consq	VEP gene
1:64951326:A:G	1	4	1:64951326:A:G	0.94	2.16×10^{-9}	<i>JAK1</i>	2.3	intron	<i>JAK1</i>
1:155197995:A:G	1	1	1:155036692:C:A	0.97	1.74×10^{-9}	<i>DCST1</i>	3.3	intron	<i>DCST1</i>
1:155197995:A:G	2	3	1:155197995:A:G	0.56	1.41×10^{-16}	<i>MUC1</i>	2.1	intron	<i>THBS3</i>
1:155197995:A:G	3	2	1:155066988:C:T	0.86	6.58×10^{-17}	<i>EFNA4</i>	10	synonymous	<i>EFNA4</i>
1:155197995:A:G	4	33	1:155278322:A:T	0.45	4.58×10^{-4}	<i>FDPS</i>	5.2	intron	<i>HCN3</i>
1:155197995:A:G	5	5	1:155302683:C:T:C	0.45	6.12×10^{-8}	<i>RUSC1</i>	0.72	upstream	<i>PKLR</i>
2:60480453:A:G	1	11	2:60480453:A:G	0.18	1.86×10^{-7}	<i>BCL11A</i>	10	intron	<i>BCL11A</i>
3:45818159:G:A	1	1	3:45329214:G:A	1	1.36×10^{-62}	<i>LARS2</i>	0.34	-	-
3:45818159:G:A	2	1	3:45796521:G:T	1	4.69×10^{-18}	<i>FYCO1</i>	9.2	5'UTR	<i>SLC6A20</i>
3:45818159:G:A	3	1	3:45239823:G:A	1	7.38×10^{-1}	<i>TMEM158</i>	0.68	-	-
3:45818159:G:A	4	1	3:45280485:C:T	1	5.03×10^{-1}	<i>TMEM158</i>	7.6	-	-
3:45818159:G:A	5	13	3:46193413:A:T	0.077	6.71×10^{-1}	<i>CCR3</i>	1.7	-	-
3:45818159:G:A	6	2	3:45696232:C:T	0.5	4.65×10^{-1}	<i>SACM1L</i>	1.9	intron	<i>SACM1L</i>
3:45818159:G:A	7	2	3:45569212:T:C	0.5	9.21×10^{-1}	<i>LIMD1</i>	5.2	-	-
3:45818159:G:A	8	2	3:45570834:G:C	0.5	6.92×10^{-1}	<i>LIMD1</i>	11	-	-
3:45818159:G:A	9	7	3:45859597:C:T	0.74	5.90×10^{-178}	<i>CXCR6</i>	0.14	intron	-
3:146517122:G:A	1	9	3:146517122:G:A	0.18	2.28×10^{-10}	<i>PLSCR1</i>	23	missense	<i>PLSCR1</i>
3:169081115:G:GGAT	1	1	3:169515899:A:AAATT	1	6.20×10^{-7}	<i>MECOM</i>	0.67	intron	<i>MECOM</i>
3:169081115:G:GGAT	2	14	3:169521598:C:T	0.15	1.98×10^{-1}	<i>MECOM</i>	2.4	intron	<i>MECOM</i>
3:169081115:G:GGAT	3	29	3:169523892:T:TACAC	0.12	7.12×10^{-8}	<i>MECOM</i>	0.68	intron	<i>MECOM</i>
4:105897896:G:A	1	1	4:105897896:G:A	1	4.65×10^{-10}	<i>NPNT</i>	24	spliceregion	<i>NPNT</i>
6:41520640:G:A	1	8	6:41522116:T:C	0.25	1.45×10^{-7}	<i>FOXP4</i>	5.2	downstream	<i>LINC01276</i>
7:100032719:C:T	1	4	7:100032719:C:T	0.48	7.96×10^{-8}	<i>TRIM4</i>	1.8	intron	<i>ZKSCAN1</i>
9:21206606:C:G	1	2	9:21328701:G:A	0.53	3.38×10^{-8}	<i>KLHL9</i>	11	downstream	<i>KLHL9</i>
9:21206606:C:G	2	3	9:21206606:C:G	0.77	1.11×10^{-11}	<i>IFNA10</i>	24	missense	<i>IFNA10</i>
9:133271182:T:C	1	11	9:133271182:T:C	0.79	1.32×10^{-8}	-	4.4	intron	<i>ABO</i>
11:34482745:G:A	1	3	11:34482745:G:A	0.58	1.62×10^{-14}	<i>CAT</i>	0.073	intron	<i>ELF5</i>
12:112919388:G:A	1	64	12:112919388:G:A	0.11	5.02×10^{-9}	<i>OAS1</i>	4.2	spliceacceptor	<i>OAS1</i>
12:132481571:G:A	1	1	12:132158547:G:A	1	3.86×10^{-12}	<i>DDX51</i>	0.31	-	-
12:132481571:G:A	2	12	12:132537984:G:A	0.19	2.70×10^{-11}	<i>FBRSL1</i>	2.9	intron	<i>FBRSL1</i>
12:132481571:G:A	3	12	12:132536375:T:C	0.28	1.76×10^{-3}	<i>POLE</i>	4	intron	<i>FBRSL1</i>
13:112881427:C:T	1	4	13:112882313:A:G	0.31	1.49×10^{-13}	<i>ATP11A</i>	0.1	3'UTR	<i>ATP11A</i>
15:31319838:G:T	1	1	15:31334071:C:T	1	1.34×10^{-8}	<i>KLF13</i>	5.6	intron	<i>KLF13</i>
16:29843264:G:A	1	12	16:29881840:G:A	0.08	4.16×10^{-6}	<i>SEZ6L2</i>	2	intron	<i>SEZ6L2</i>
16:89196249:G:A	1	2	16:89196249:G:A	0.74	4.67×10^{-10}	<i>SLC22A31</i>	23	missense	<i>SLC22A31</i>
17:49863260:C:A	1	1	17:50213197:T:C	1	2.16×10^{-9}	<i>TMEM92</i>	2.7	intron	-
17:49863260:C:A	2	1	17:50030360:G:A	0.99	1.58×10^{-7}	<i>ITGA3</i>	0.29	-	-
17:49863260:C:A	3	1	17:50109066:C:T	1	6.74×10^{-1}	<i>ITGA3</i>	2.5	downstream	<i>SAMD14</i>
17:49863260:C:A	4	2	17:50212932:G:A	0.94	5.09×10^{-1}	<i>TMEM92</i>	1.6	intron	-
17:49863260:C:A	5	2	17:50225230:G:T	0.7	7.19×10^{-1}	<i>COL1A1</i>	1	-	-
17:49863260:C:A	6	3	17:49863260:C:A	0.69	3.85×10^{-11}	<i>DLX3</i>	5.4	-	-
19:4717660:A:G	1	1	19:4717660:A:G	0.99	1.40×10^{-46}	<i>DPP9</i>	16	intron	<i>DPP9</i>
19:4717660:A:G	2	1	19:5066102:C:T	1	1.87×10^{-14}	<i>KDM4B</i>	0.2	intron	<i>KDM4B</i>
19:10352442:G:C	1	1	19:10352442:G:C	0.98	2.31×10^{-25}	<i>TYK2</i>	25	missense	<i>TYK2</i>
19:10352442:G:C	2	1	19:10414696:G:A	1	1.85×10^{-11}	<i>ICAM1</i>	1.1	-	-
19:10352442:G:C	3	1	19:10020130:A:C	0.95	8.14×10^{-7}	<i>RDH8</i>	2.8	intron	<i>RDH8</i>
19:48702888:G:C	1	1	19:48496300:TC:T	1	9.88×10^{-14}	<i>LMTK3</i>	0.78	intron	<i>LMTK3</i>
19:48702888:G:C	2	1	19:48500758:G:A	1	2.31×10^{-1}	<i>LMTK3</i>	8	intron	<i>LMTK3</i>
19:48702888:G:C	3	5	19:48466475:C:T	0.25	6.18×10^{-1}	<i>CYTH2</i>	0.14	3'UTR	<i>KCNJ14</i>
19:48702888:G:C	4	9	19:48697960:C:T	0.52	5.23×10^{-13}	<i>RASIP1</i>	2.4	intron	<i>FUT2</i>
21:33240996:A:G	1	1	21:33412149:C:T	1	1.53×10^{-37}	<i>IFNAR2</i>	2.2	intron	<i>IFNGR2</i>
21:33240996:A:G	2	1	21:34355721:GGAAA:G	1	1.29×10^{-33}	<i>ATP5PO</i>	0.63	intron	-
21:33240996:A:G	3	1	21:33441927:G:A	1	1.97×10^{-2}	<i>IFNAR2</i>	0.48	downstream	<i>IFNGR2</i>
21:33240996:A:G	4	18	21:33232252:G:A	0.094	1.59×10^{-25}	<i>IL10RB</i>	0.43	intron	<i>IFNAR2</i>
21:33240996:A:G	5	14	21:33954962:A:G	0.15	8.39×10^{-15}	<i>MRPS6</i>	3.8	intron	<i>LINC00649</i>
21:33240996:A:G	6	31	21:33287378:C:T	0.097	6.15×10^{-11}	<i>IFNAR1</i>	3.6	intron	<i>IL10RB</i>
X:15602217:T:C	1	1	X:15752522:A:G	1	6.95×10^{-12}	<i>CA5B</i>	2.6	intron	<i>CA5B</i>
X:15602217:T:C	2	1	X:15644354:C:T	1	5.08×10^{-9}	<i>CLTRN</i>	2.5	downstream	-

Table 7: Lead variants in credible sets from **European (EUR)** fine-mapping in the GenOMICC severe versus ctrl-all cohort. Locus index variant indicates index variant for locus as identified by the multi-study meta-analysis and is the variant with the highest posterior inclusion probability (PIP) in each credible set. *iCS* and *nCS* indicate the index and the number of variants in each credible set. *PIP* is the posterior inclusion probability of the lead variant. *P* indicates the *P*-value in the EUR GWAS for the lead variant. *V2G gene* is the gene prioritised with the highest score by the Variant-to-Gene (V2G) pipeline on the Open Targets Genetics platform for the lead variant. *CADD* is the Combined Annotation Dependent Depletion score for the lead variant while *VEP consq* and *VEP gene* indicate the VEP predicted consequence on the canonical transcript and impacted gene, respectively.¹³

Locus index variant	iCS	nCS	CS worst variant	PIP	P	CADD	VEP consq	VEP gene
1:155197995:A:G	2	3	1:155202934:T:C	0.29	5.34×10^{-16}	21	missense	<i>THBS3</i>
1:155197995:A:G	4	33	1:155260340:C:T	0.009	8.14×10^{-3}	2.5	synonymous	<i>SCAMP3</i>
9:133271182:T:C	1	11	9:133257521:T:TC	0.0072	2.00×10^{-6}	15	frameshift	<i>ABO</i>
19:48702888:G:C	4	9	19:48703346:C:T	0.1	3.41×10^{-12}	7	synonymous	<i>FUT2</i>
21:33240996:A:G	4	18	21:33241950:T:G	0.057	2.66×10^{-25}	0.002	missense	<i>IFNAR2</i>

Table 8: Variants with worst consequence in credible sets from **European (EUR)** fine-mapping in the GenOMICC severe versus ctrl-all cohort. Variants with the worst VEP consequence in each fine-mapped credible are shown, for cases when the worst variant is not the lead variant in terms of PIP in the credible set and the consequence has low impact or worse (“LOW”, “MODERATE” or “HIGH” in VEP impact scale). *PIP*, *P*, *CADD*, *VEP consq* and *VEP gene* columns refer to the CS worst variant. *P* indicates *P*-values in the EUR GWAS. *CADD* is the Combined Annotation Dependent Depletion score while *VEP consq* and *VEP gene* indicate the VEP predicted consequence on the canonical transcript and impacted gene, respectively.

1.1.4 Age and Sex Stratified analysis

We performed an age-stratified analysis for European ancestry individuals dividing them between < 60 and ≥ 60 years old (the median age for cases; Table 1). We also performed a sex-stratified analysis for European ancestry individuals. in both cases, we compared the two cohorts using the same methodology as¹². We did not find genome wide significant differences in effects for either analysis. In Fig. 5 we show a comparison of the effects at the lead variants of our main analysis.

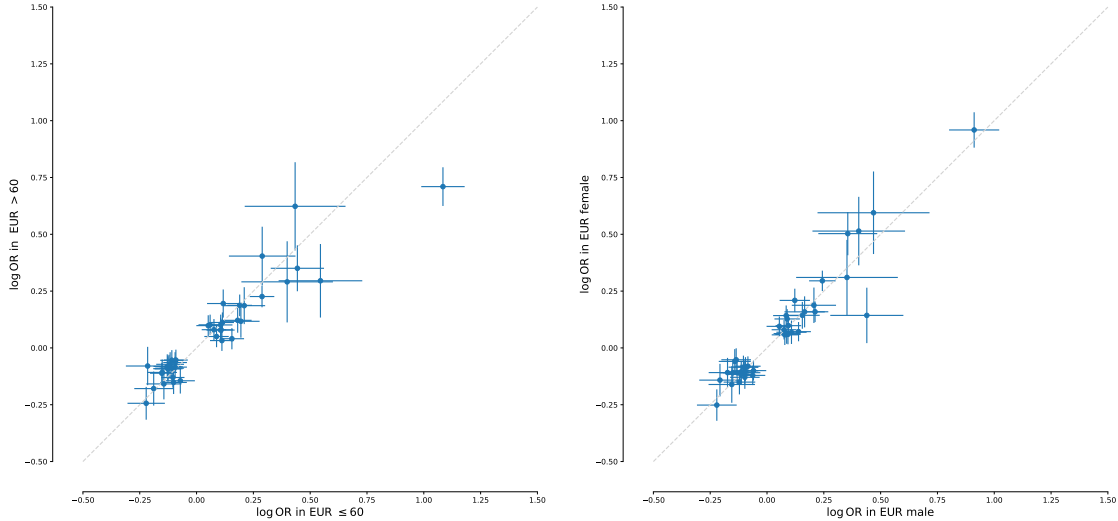


Figure 5: Comparison of logOR estimates with 95% CI at lead variants of the main analysis for age and sex stratified GWAS in European ancestry individuals.

1.2 Polygenic risk score (PRS) analysis

With 80% of the critical cases vs. ctrl-mld cohort we calculated a polygenic risk score (PRS) for individuals of European ancestry. We then tested the PRS on the remaining 20% of individuals from European ancestry. We compared the genetic PRS to a risk score based on covariates comprising sex and age. A model including PRS and covariates improved prediction of Covid-19 cases ($\text{auc}=0.73$) compared with a model using just covariates ($\text{auc}=0.70$). The mean age of the top 5% individuals with critical Covid-19 is 58.83 and 62% of them are males compared to a mean age of 60.1 and a percentage of 66% of males in our test dataset.

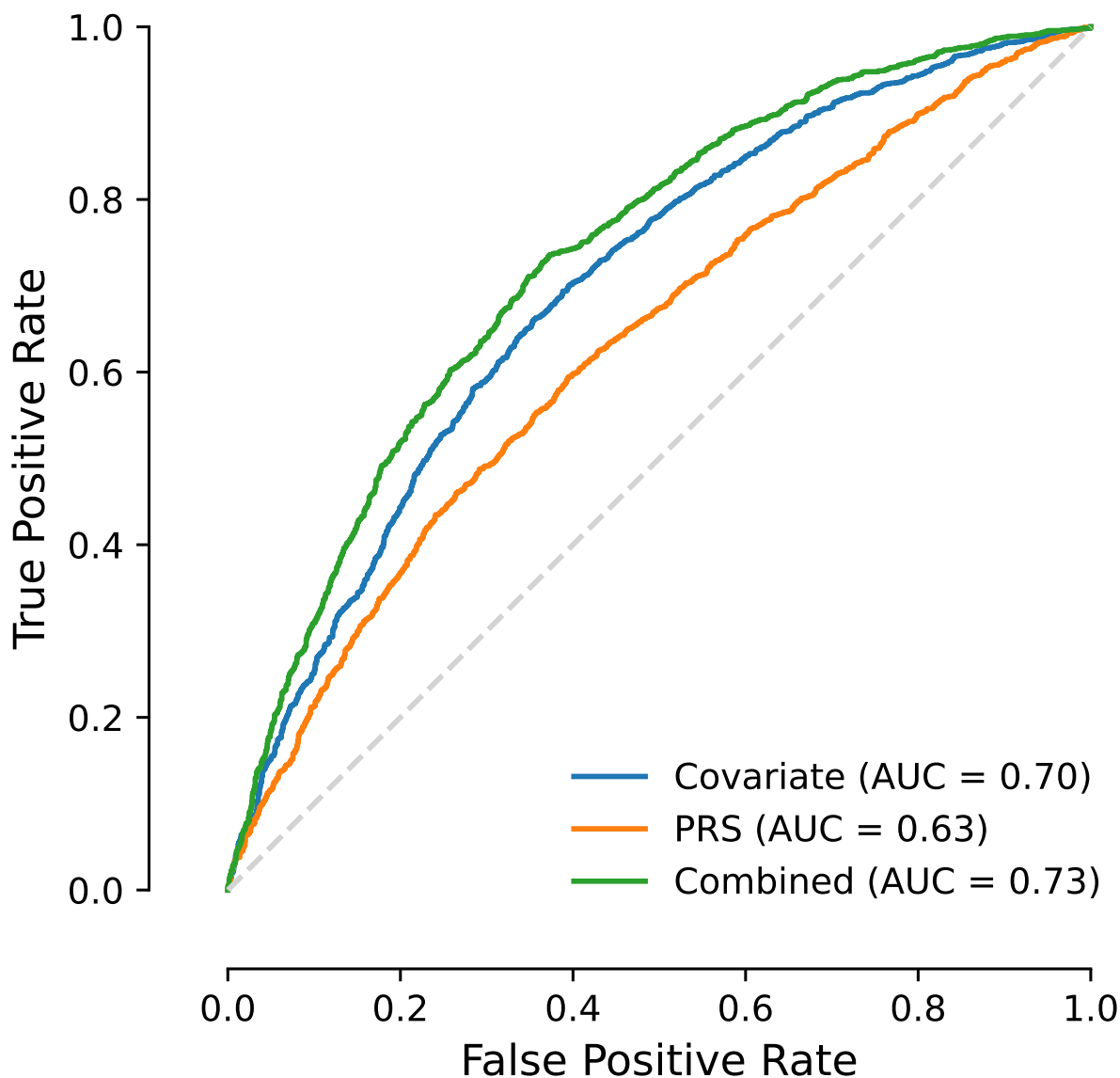


Figure 6: Receiver operator curves for risk scores on validation cohort of European ancestry, comparing a covariate (Covariate) and genetic (PRS) risk scores, as well as their sum (Combined).

1.3 Gene-based rare variant analysis tests (RVAT)

1.3.1 RVAT Cohorts

For the gene-based rare variant association analysis, we used a subset of 28,091 individuals from our cohorts, consisting of 11,423 individuals with critical COVID-19 from GenOMICC and 16,668 control individuals. The control group included 11,268 individuals from the mild COVID-19 cohort and 5,400 from the 100kGP cohort, all processed with the same alignment and variant calling pipeline (a control set referred to as ctrl-dgn; Extended Data Fig.1; Methods). The reasoning for using the ctrl-dgn cohort for the RVAT analysis is as follows: (1) We used a subset of controls processed with the same pipeline as opposed to using all available controls (ctrl-all) to minimise potential batch effects as those would be more difficult to quality control than our GWAS analysis on common variants. (2) We included 100kGP individuals in the controls to increase discovery power, boosting control numbers by $\sim 50\%$ compared to using only ctrl-mld.

Six additional studies contributed data for our multi-study RVAT meta-analysis. Case/control breakdowns across all contributing studies are shown in Extended Data Fig.1 and Table 9.

Study cohort	Ancestry	N_{cases}	$N_{controls}$
GenOMICC	EUR	9,128	14,670
GenOMICC	SAS	1,229	1,165
GenOMICC	AFR	681	594
GenOMICC	EAS	385	239
UK Biobank	EUR	2,101	465,275
GENCOVID	EUR	1,362	705
BQC19-SweCovid	EUR	361	1,143
DeCOI	EUR	243	587
POLCOVID	EUR	109	1249
PMBB	AFR	107	10,264
PMBB	EUR	80	28,964
Total	-	15,786	524,855

Table 9: Sample sizes for each study cohort included in the multi-study RVAT meta-analysis.

1.3.2 Multi-study RVAT meta-analysis

To assess the role of rare variants in critical illness, we conducted gene-based analysis with REGENIE¹³. We defined two sets of variant masks: *strict-LoF* including variants with predicted loss of function effects

(LoF) and *mild_LoF*, including variants annotated as either LoF or missense (Methods). To maximise power, we combined strength across multiple variant tests (Burden¹⁴, SKAT-O¹⁵, ACAT-V¹⁶) and allele frequency thresholds (singletons, AF=0.005) with the Cauchy combination test and calculated a single P -value (ACAT-O; Methods). We then meta-analysed gene-wise ACAT-O P -values across all studies (Table 9) using Stouffer’s method and weighted by effective sample size (n_w ; Methods). We set a Bonferroni corrected gene-wide significance threshold of 2.58×10^{-6} (0.05 divided by 19,357 protein coding genes tested).

From the multi-study RVAT meta-analysis, we obtained *TLR7* as gene-wide significant for the *strict_LoF* mask (P -value= 8.25×10^{-8} ; Figure 7A) and *SLC50A1* and *TLR7* as gene-wide significant for the *mild_LoF* mask (P -value = 1.04×10^{-6} and 1.74×10^{-6} , respectively; Figure 7B). When considering contributions from each study, *SLC50A1* was nominally significant ($P < 0.05$) in three studies (GenOMICC, UK Biobank and BQC19-SweCovid), while *TLR7* was nominally significant in two studies (GenOMICC and UK Biobank), Table 10.

To replicate the novel *SLC50A1* signal for the *mild_LoF* mask, we performed a leave-largest-study-out meta-analysis, excluding the GenOMICC cohorts and meta-analysing the remaining studies obtaining a nominally significant P -value ($P=0.0289$) for *SLC50A1*.

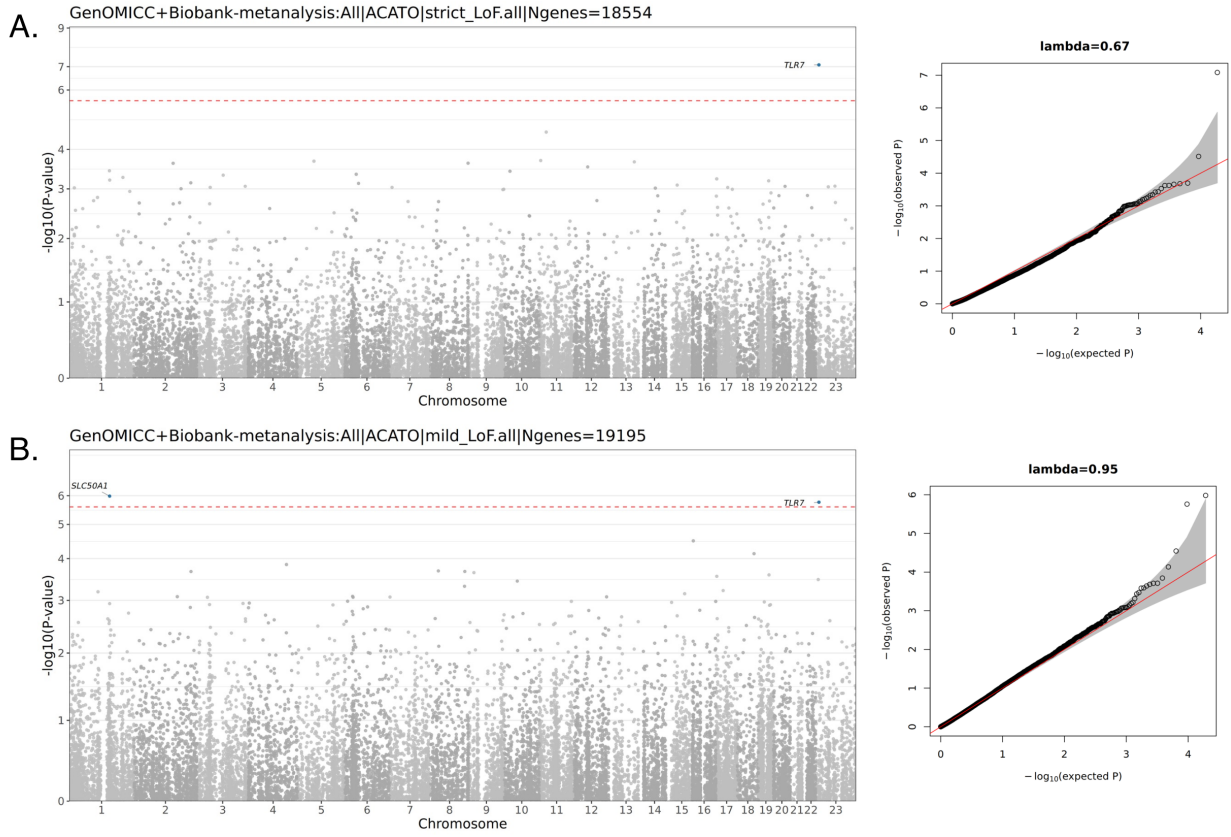


Figure 7: Results from rare variant multi-study meta-analysis. A. Results for *strict_lof* mask. B. Results for *mild_lof* mask. Genes that are gene-wide significant at Bonferroni corrected threshold $P < 2.58 \times 10^{-6}$ (red dashed line) are highlighted in blue. ACAT-O gene-based P -values are shown.

Cohort	Gene	Mask	P	OR CI burden	Best test	Best test P
GenOMICC-EUR	<i>SLC50A1</i>	mild-LoF	1.53×10^{-5} *	1.91[1.42-2.56]	SKATO	1.21×10^{-5}
GenOMICC-EUR	<i>SLC50A1</i>	strict-LoF	4.11×10^{-1}	2.1[0.359-12.2]	ACATV	4.11×10^{-1}
GenOMICC-EUR	<i>TLR7</i>	mild-LoF	3.99×10^{-7} *	1.77[1.36-2.3]	SKATO	2.81×10^{-7}
GenOMICC-EUR	<i>TLR7</i>	strict-LoF	5.32×10^{-6} *	8.7[3.43-22.1]	ACATV	5.32×10^{-6}
GenOMICC-SAS	<i>SLC50A1</i>	mild-LoF	7.16×10^{-2}	12.6[0.801-199]	ACATV	7.16×10^{-2}
GenOMICC-SAS	<i>TLR7</i>	mild-LoF	2.68×10^{-1}	1.48[0.741-2.94]	ACATV	2.68×10^{-1}
GenOMICC-SAS	<i>TLR7</i>	strict-LoF	1.22×10^{-1}	3.27[0.729-14.7]	ACATV	1.22×10^{-1}
GenOMICC-AFR	<i>SLC50A1</i>	mild-LoF	2.55×10^{-1}	5.31[0.299-94.3]	ACATV	2.55×10^{-1}
GenOMICC-AFR	<i>SLC50A1</i>	strict-LoF	4.35×10^{-1}	5.04[0.0865-294]	ACATV	4.35×10^{-1}
GenOMICC-AFR	<i>TLR7</i>	mild-LoF	2.42×10^{-1}	2.26[0.576-8.87]	ACATV	2.42×10^{-1}
GenOMICC-AFR	<i>TLR7</i>	strict-LoF	5.73×10^{-1}	4.01[0.0319-505]	ACATV	5.73×10^{-1}
GenOMICC-EAS	<i>SLC50A1</i>	mild-LoF	1.65×10^{-1}	0.0506[0.000746-3.43]	ACATV	1.65×10^{-1}
GenOMICC-EAS	<i>TLR7</i>	mild-LoF	6.22×10^{-1}	3.62[0.0215-611]	ACATV	6.22×10^{-1}
DeCOI	<i>SLC50A1</i>	mild-LoF	8.35×10^{-1}	0.838[0.159-4.43]	ACATV	8.35×10^{-1}
DeCOI	<i>TLR7</i>	mild-LoF	3.18×10^{-1}	0.471[0.107-2.06]	ACATV	3.18×10^{-1}
GENCOVID	<i>SLC50A1</i>	mild-LoF	6.52×10^{-1}	1.31[0.409-4.17]	ACATV	6.52×10^{-1}
GENCOVID	<i>TLR7</i>	mild-LoF	4.47×10^{-1}	1.57[0.49-5.05]	ACATV	4.47×10^{-1}
PMBB-EUR	<i>SLC50A1</i>	mild-LoF	2.57×10^{-1}	2.94[0.171-50.5]	ACATV	2.45×10^{-1}
PMBB-EUR	<i>SLC50A1</i>	strict-LoF	9.60×10^{-1}	0.367[5.04e-18-2.67e+16]	ACATV	9.60×10^{-1}
PMBB-AFR	<i>SLC50A1</i>	mild-LoF	6.99×10^{-1}	0.35[0.0175-7]	FIRTH	4.92×10^{-1}
UK Biobank	<i>SLC50A1</i>	mild-LoF	2.81×10^{-2} *	1.19[0.685-2.07]	SKATO	2.56×10^{-2}
UK Biobank	<i>SLC50A1</i>	strict-LoF	5.51×10^{-1}	0.36[0.0381-3.4]	FIRTH	3.73×10^{-1}
UK Biobank	<i>TLR7</i>	mild-LoF	4.21×10^{-1}	1.01[0.945-1.07]	ACATV	2.09×10^{-1}
UK Biobank	<i>TLR7</i>	strict-LoF	1.50×10^{-3} *	13.9[2.74-70.8]	ACATV	1.50×10^{-3}
BQC19-SweCovid	<i>SLC50A1</i>	mild-LoF	4.05×10^{-2} *	5.13[1.07-24.5]	ACATV	4.05×10^{-2}
BQC19-SweCovid	<i>SLC50A1</i>	strict-LoF	6.52×10^{-1}	0.295[0.00146-59.4]	ACATV	6.52×10^{-1}
BQC19-SweCovid	<i>TLR7</i>	mild-LoF	9.96×10^{-1}	0.995[0.17-5.83]	ACATV	9.96×10^{-1}

Table 10: Results for significant genes for RVAT analysis across studies. *P* indicates ACAT-O gene-based *P*-values with asterisk (*) indicating nominally significant *P*-value ($P < 0.05$) in the respective cohort and mask. *OR CI burden* shows the odds ratio and 95% Confidence interval (CI) for the burden test, *Best test* shows the test that produced the minimum *P*-value among the combined tests FIRTH (burden), SKATO, ACATV within ACAT-O test. *Best test P* shows the *P*-value of the displayed *best test*. Cohorts and mask combinations that yielded no variants in the respective genes after applying the respective mask are not shown.

1.3.3 Individual variants contributing to the RVAT gene signals

We examined individual variant contributions from the EUR GenOMICC vs ctrl-dgn analysis to the *SLC50A1* and *TLR7* signals to understand the genetic architecture of the association signals by estimating their effect sizes with Firth’s logistic regression as implemented in REGENIE. Gene *SLC50A1* had 24 missense and 3 LoF mutations. We identified *SLC50A1* missense variant 1:155138217:G:T with minor allele frequency (maf)=0.00242 (P -value = 2.03×10^{-5} , OR[95%] = 2.31[1.57-3.39], CADD=28.6) as the only contributing

variant with P -value < 0.05 . Gene *TLR7* had 11 LoF and 45 missense mutations and there were no variants with minor allele count (MAC) ≥ 10 with P -values < 0.05 . Two *TLR7* variants with MAC < 10 , LoF variant X:12885886:CACTT:C (P -value $= 1.9 \times 10^{-2}$, OR[95%] $= 5.05[1.31-19.6]$, CADD=13.2) and missense variant X:12886145:G:C (P -value $= 9.66 \times 10^{-3}$, OR[95%] $= 7.43[1.63-34]$, CADD=14.5), were nominally significant ($P < 0.05$).

1.3.4 Conditional analyses for *SLC50A1*

The *SLC50A1* gene signal was identified as gene-wide significant in the multi-study RVAT meta-analysis. The gene resides within a GWAS significant locus marked by the index variant 1:155197995:A:G (rs41264915) at 1q22 (Table 2). We conducted conditional RVAT analyses to assess whether the *SLC50A1* RVAT signal is conditionally independent of the common variant signals identified through cross-ancestry or EUR fine-mapping within that locus (Table 5, and Table 7, respectively). We used the GenOMICC severe vs ctrl-dgn EUR cohort results for these conditional analyses as it has the largest number of cases and is the greatest contributor to the RVAT signal. Across conditioned variants, the *SLC50A1* gene signal P -value was reduced in significance only when conditioning on the common variant 1:155036692:C:A that was fine-mapped in EUR (Table 11). Therefore, we hypothesise that lead variant signal 1:155036692:C:A is partially tagging the *SLC50A1* gene signal.

CondVar	Source	Pval
1:155066988:C:T	xancestry/EUR	1.12×10^{-5}
1:155197995:A:G	xancestry/EUR	1.08×10^{-5}
1:155278322:A:T	xancestry/EUR	1.06×10^{-5}
1:155175305:G:A	xancestry	1.35×10^{-5}
1:155036692:C:A	EUR	2.79×10^{-3}
1:155302683:CT:C	EUR	1.33×10^{-5}
-	-	1.53×10^{-5}

Table 11: Conditional analyses for the RVAT *SLC50A1* gene signal in EUR GenOMICC severe vs ctrl-dgn cohort for the *mild_lof* mask. *CondVar* is the conditioned lead variant from the credible sets identified in multi-ancestry fine-mapping (xancestry), EUR fine-mapping (EUR), or both (xancestry/EUR). Last row presents unconditional results. P -values shown (column: Pval) are from the ACAT-O gene-based test.

1.3.5 RVAT analysis stratified by age

To explore genetic effects for COVID-19 severity in younger versus older individuals, we conducted RVAT multi-ancestry meta-analysis on GenOMICC severe vs ctrl-dgn cohort subsets that were stratified by age. As age 60 was the median age in cases (Table 1), we split the cohorts in age <60 and age ≥ 60 to evenly split the dataset and have comparable statistical power in the two groups. The age <60 analysis re-produced the signals for the *mild_LoF* mask for both *SLC50A1* and *TLR7* (P -value $= 1.11 \times 10^{-6}$ and 8.35×10^{-8} , respectively; Figure 8A). For age ≥ 60 we observed attenuation of both the *SLC50A1* and *TLR7* signals

(Figure 8B). The *strict_LoF* mask produced a significant signal in the age<60 analysis for gene *VPS13B* (P -value = 1.78×10^{-6} ; Figure 9A) and none in the age ≥ 60 cohort (Figure 9B). As *VPS13B* was not significant in the multi-study meta-analysis (P -values of 0.85 and 0.63 for the *strict_LoF* nor *mild_LoF* masks, respectively), we did not follow it up further.

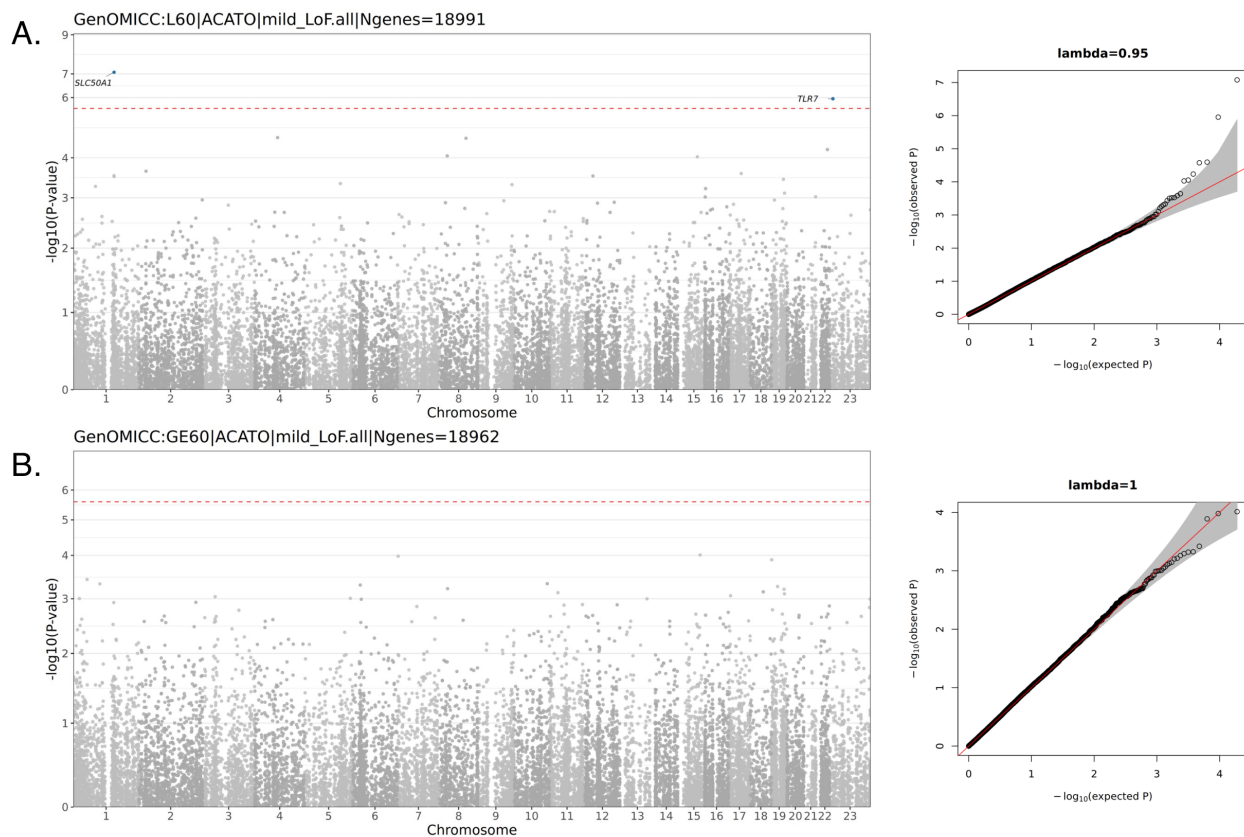


Figure 8: Results from rare variant multi-ancestry meta-analysis for cohorts stratified by age for the *mild_lof* mask. A. Results for individuals with age < 60; B. Results for individuals with age ≥ 60 . Genes that are gene-wide significant at Bonferroni corrected threshold $P < 2.58 \times 10^{-6}$ (red dashed line) are highlighted in blue. ACAT-O gene-based P -values are shown.

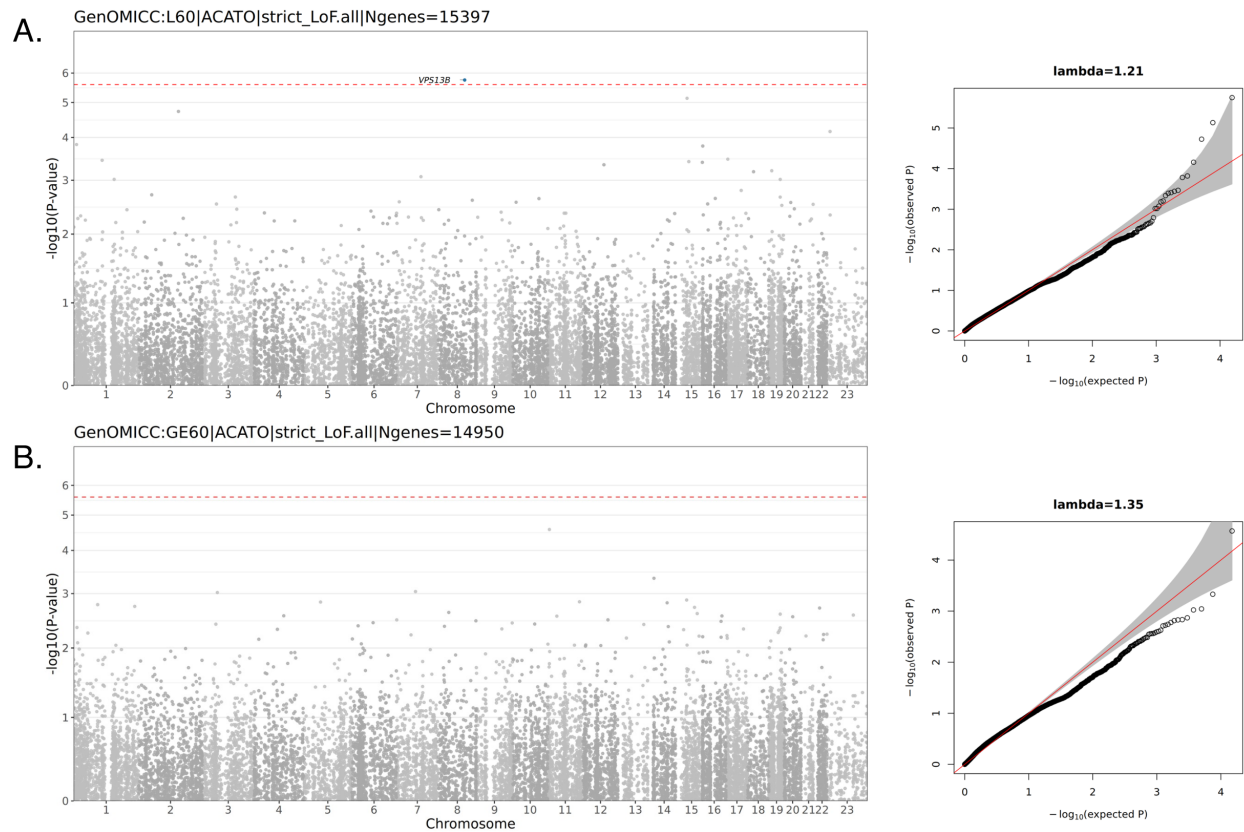
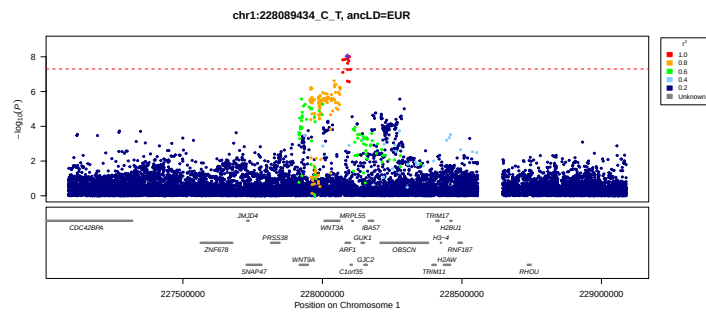
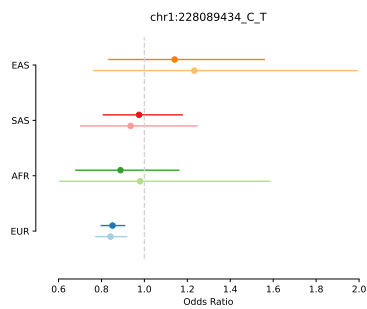
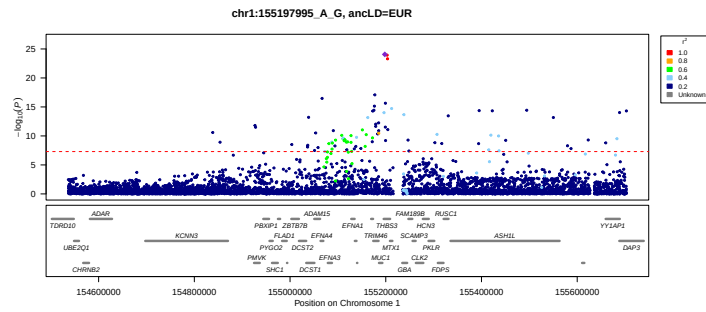
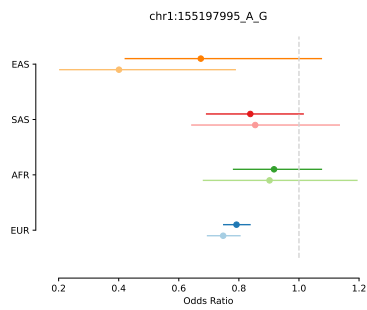
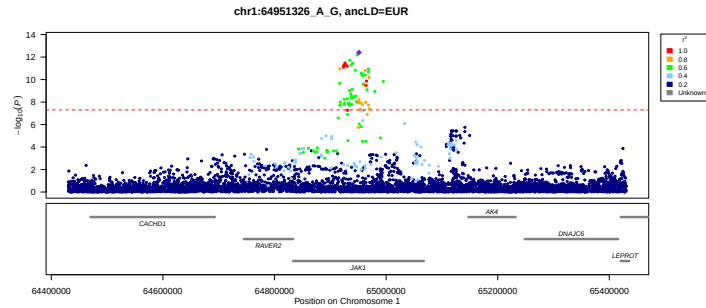
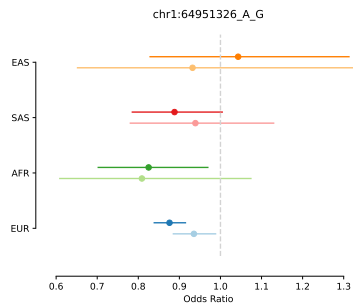
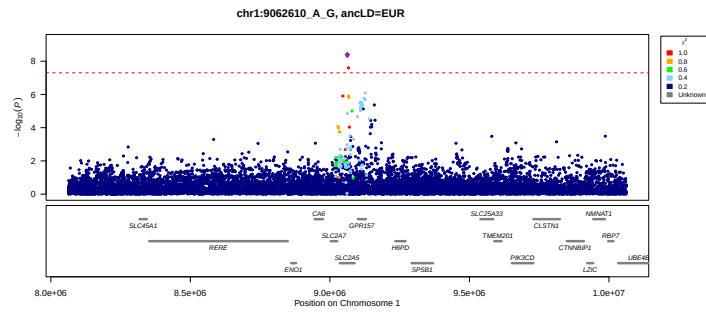
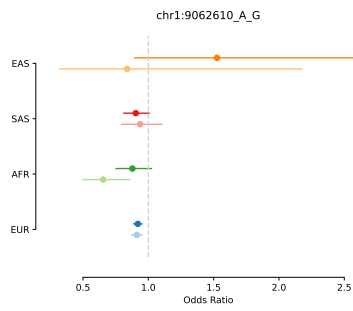
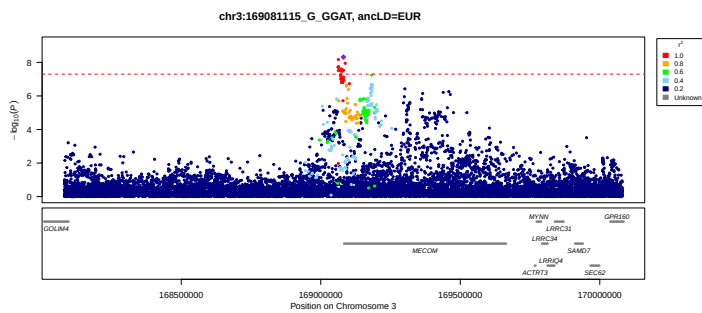
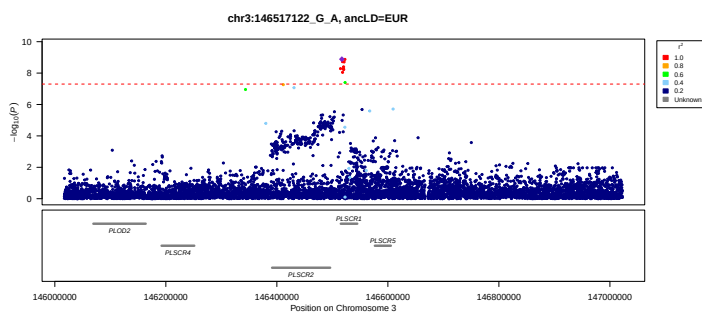
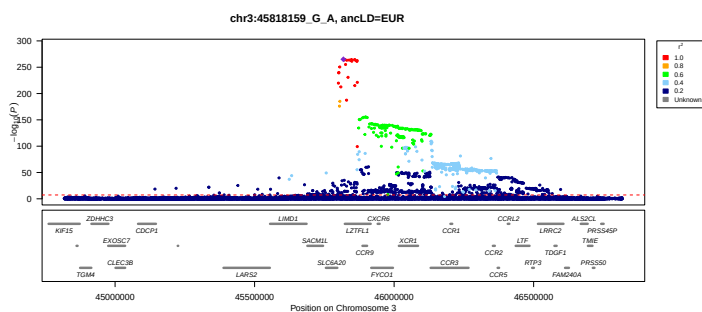
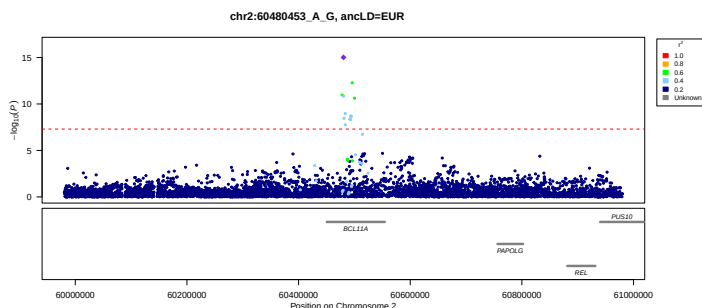
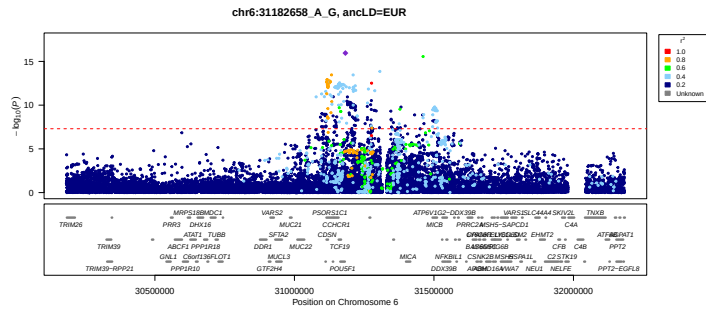
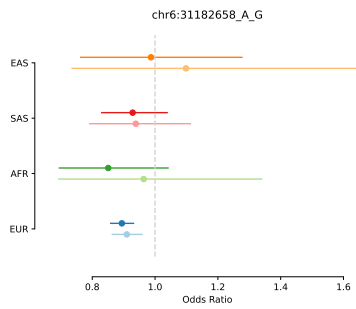
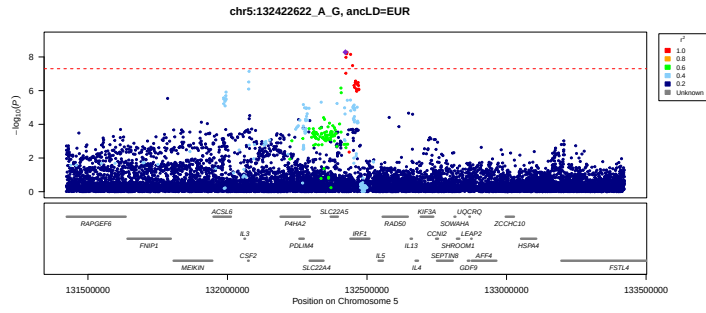
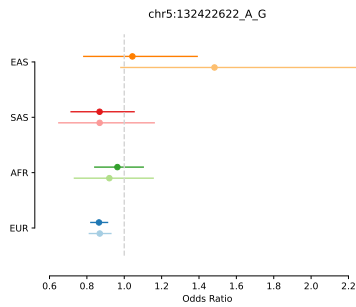
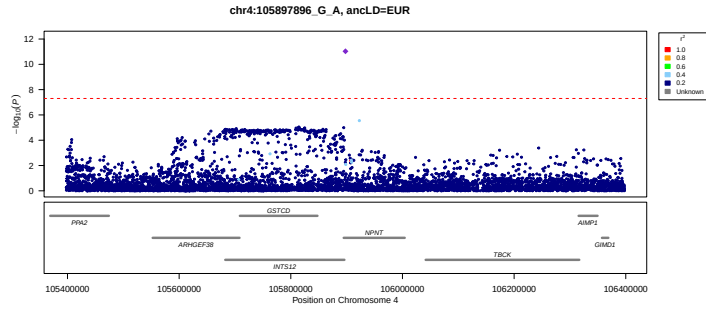
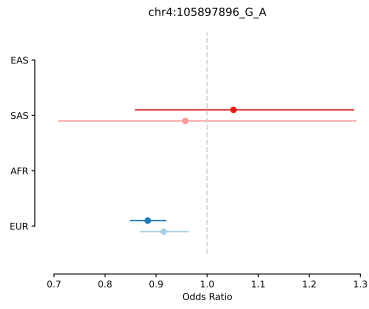
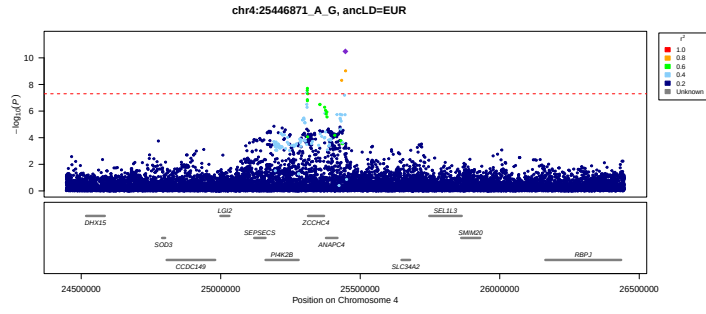
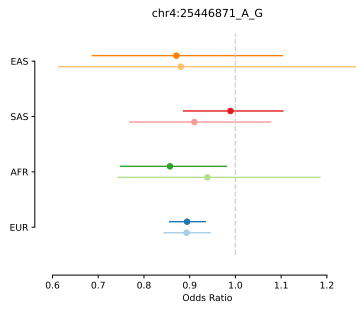


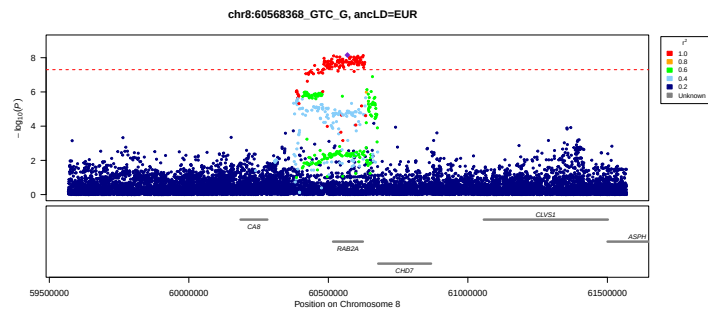
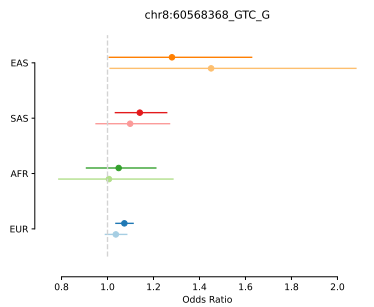
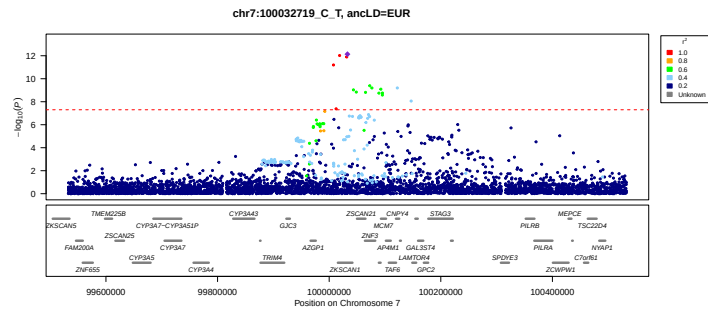
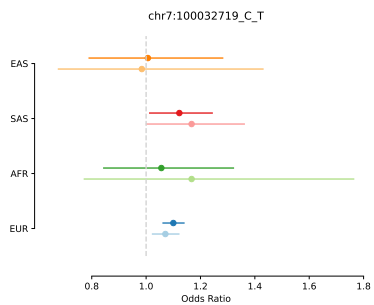
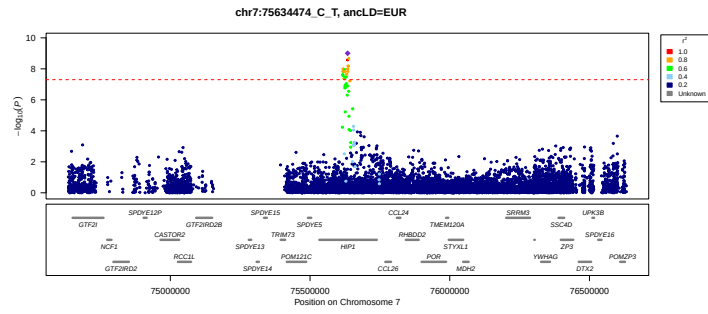
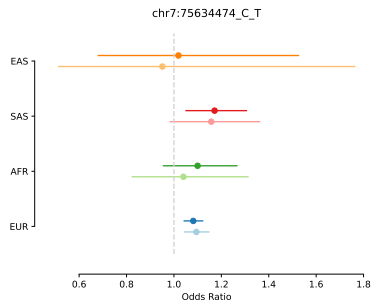
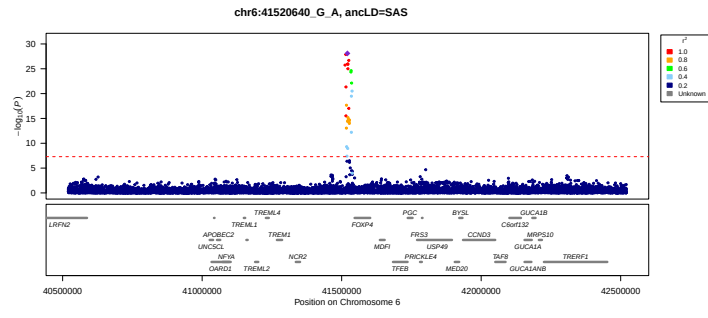
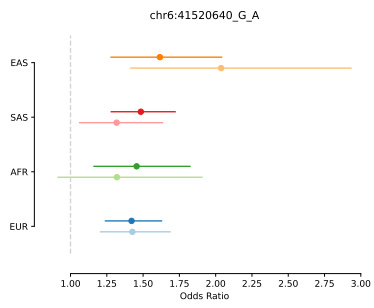
Figure 9: Results from rare variant multi-ancestry meta-analysis for cohorts stratified by age for the *strict_lof* mask. A. Results for individuals with age < 60; B. Results for individuals with age ≥ 60 . Genes that are gene-wide significant at Bonferroni corrected threshold $P < 2.58 \times 10^{-6}$ (red dashed line) are highlighted in blue. ACAT-O gene-based P -values are shown.

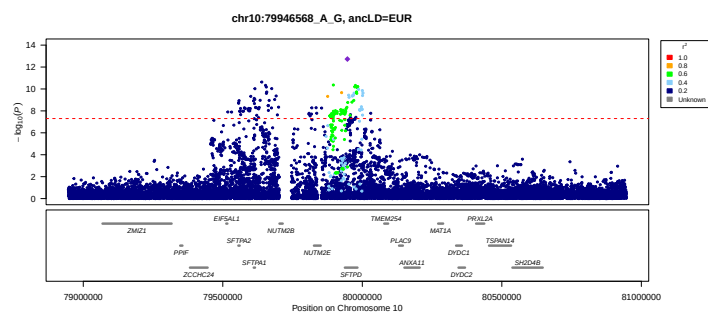
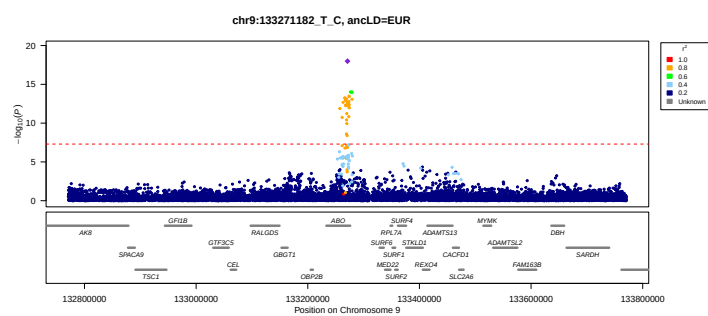
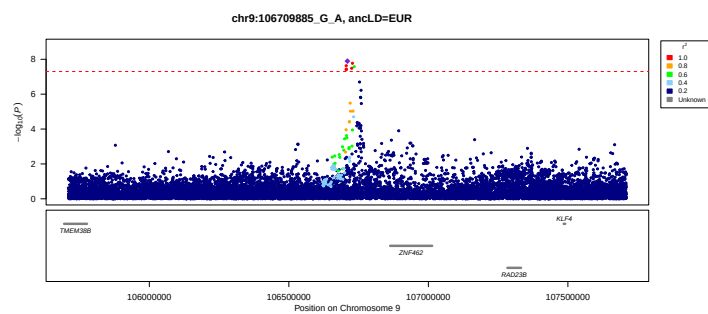
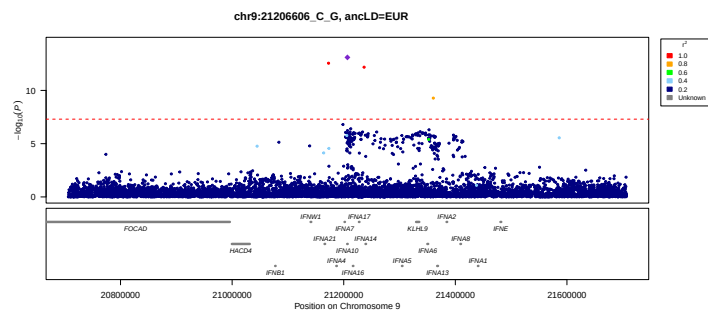
2 Additional figures

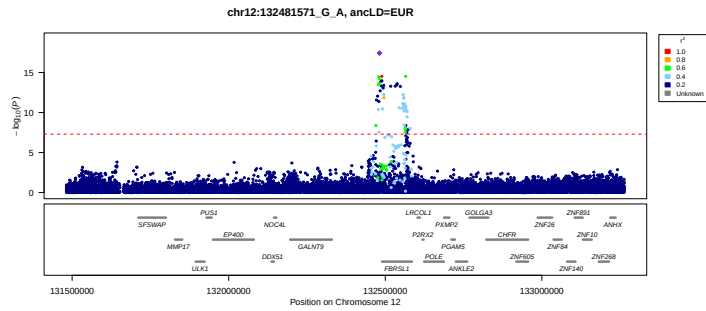
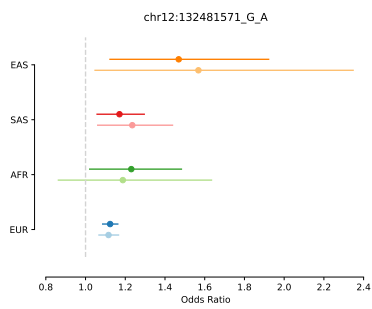
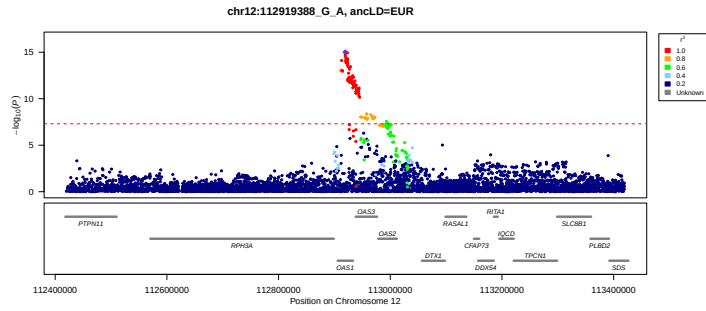
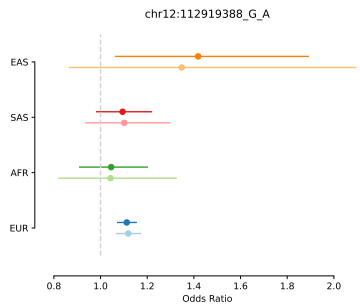
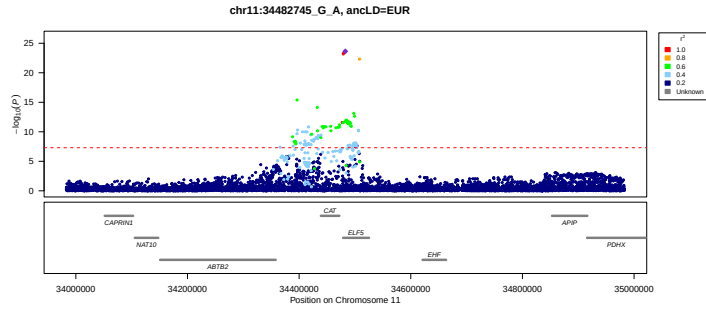
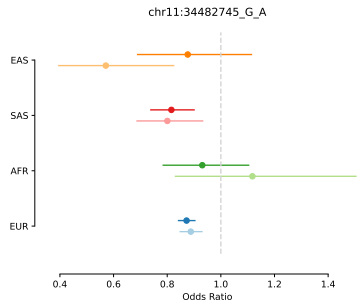
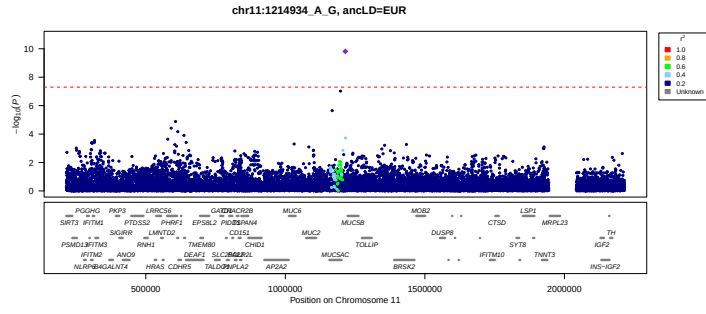
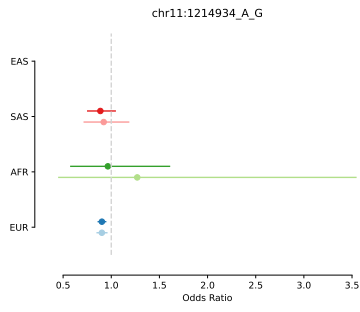


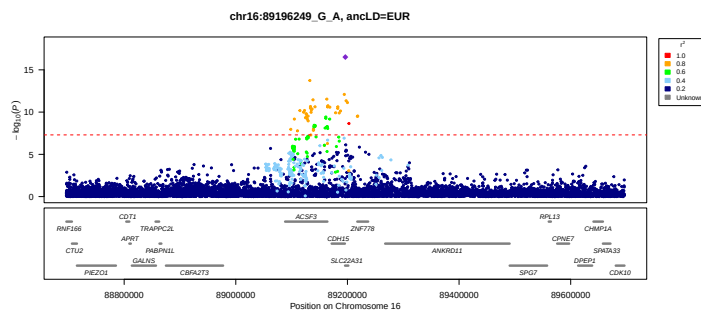
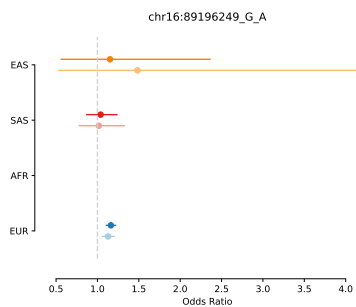
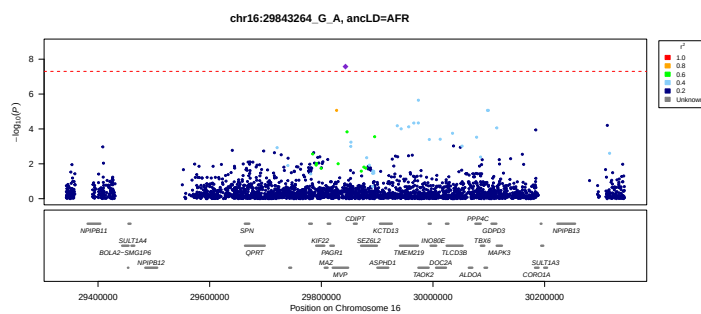
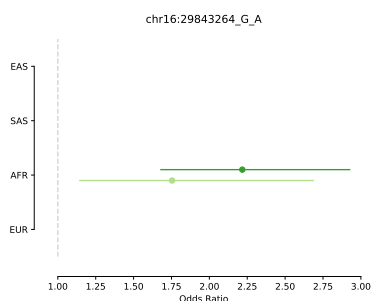
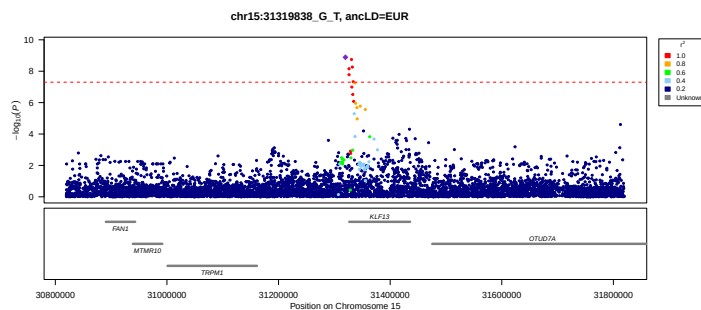
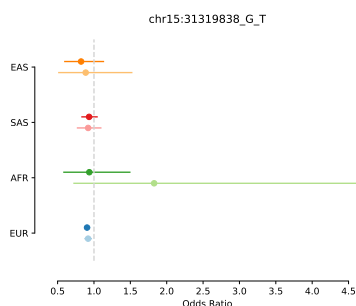
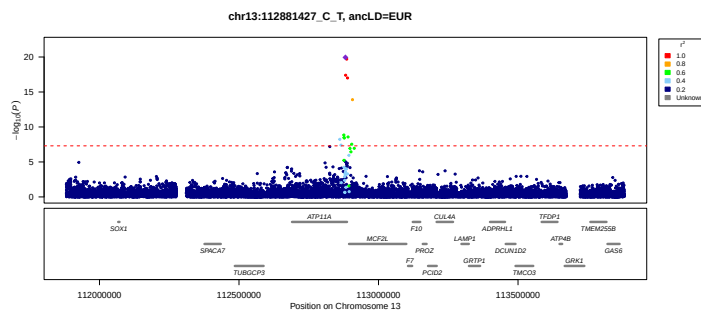
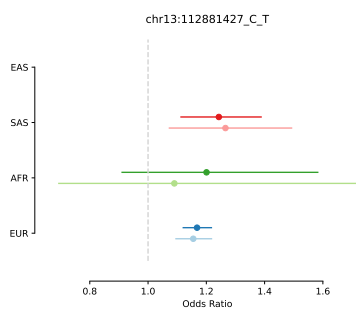


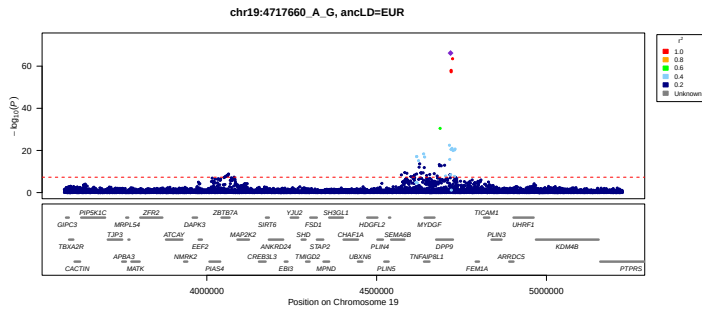
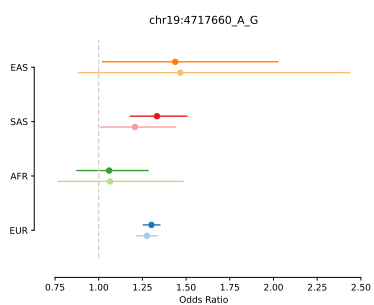
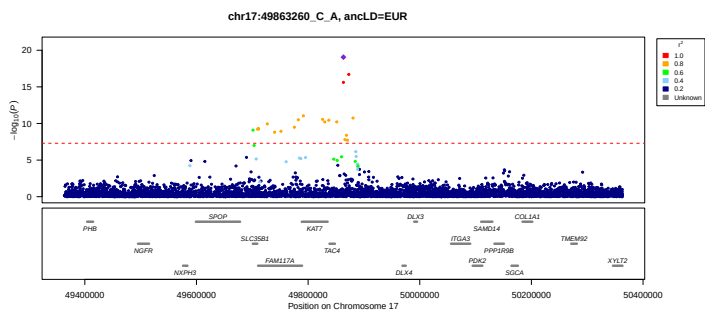
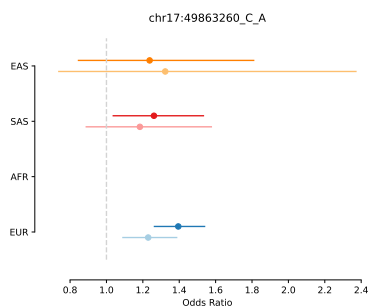
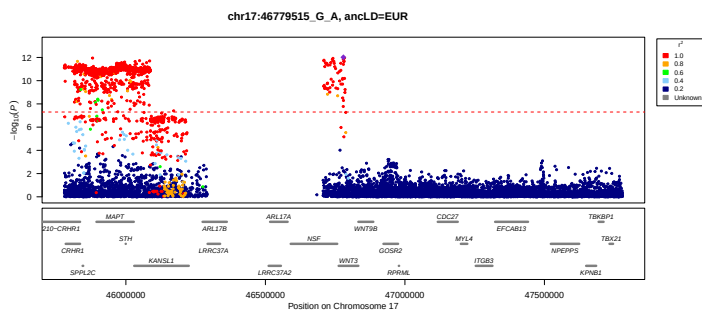
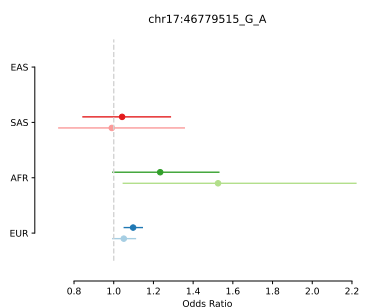
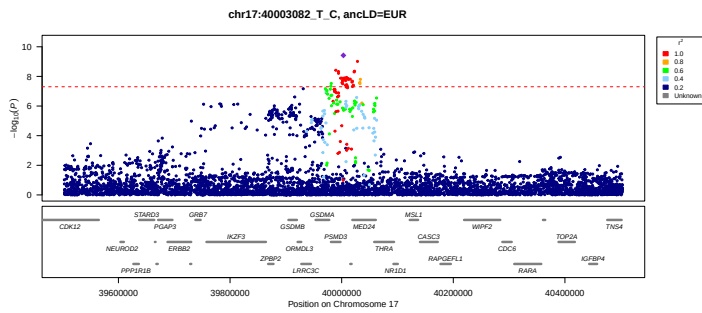
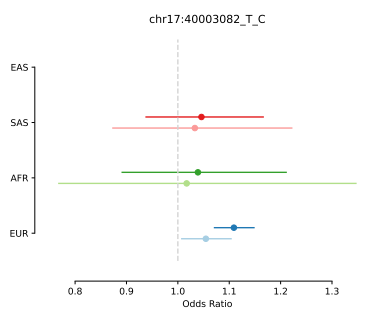


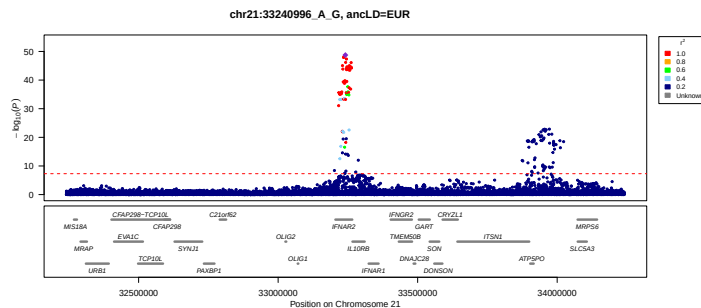
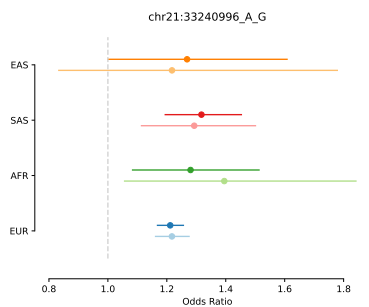
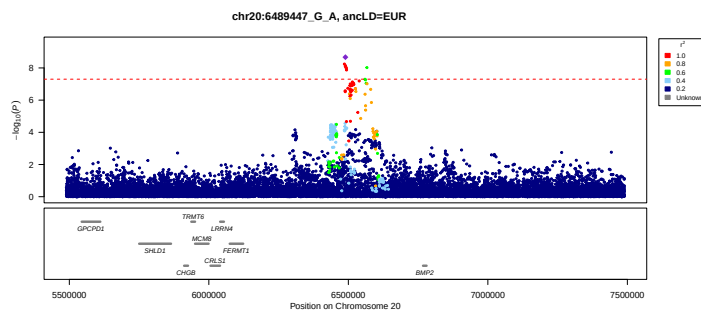
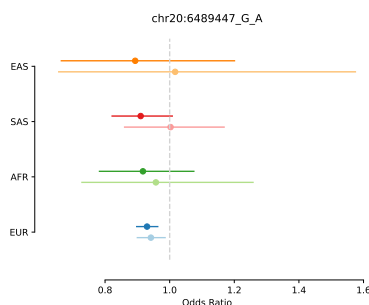
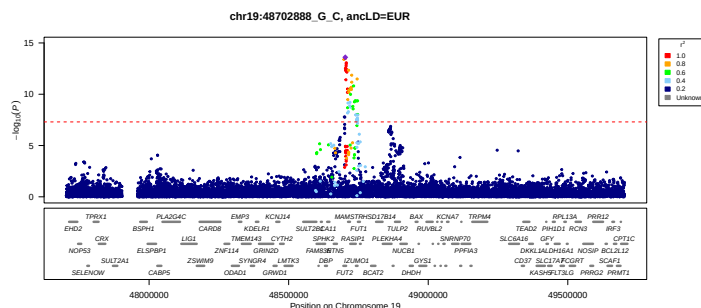
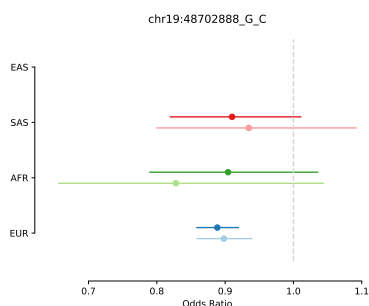
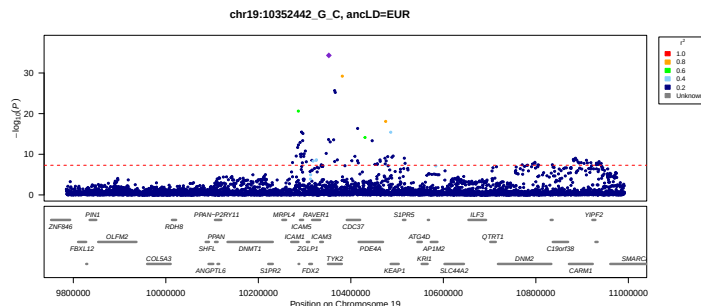
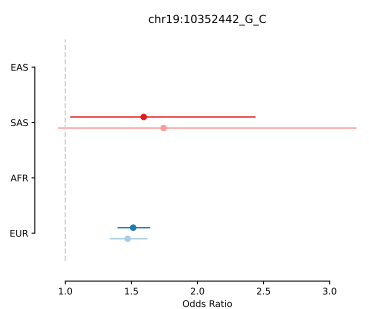












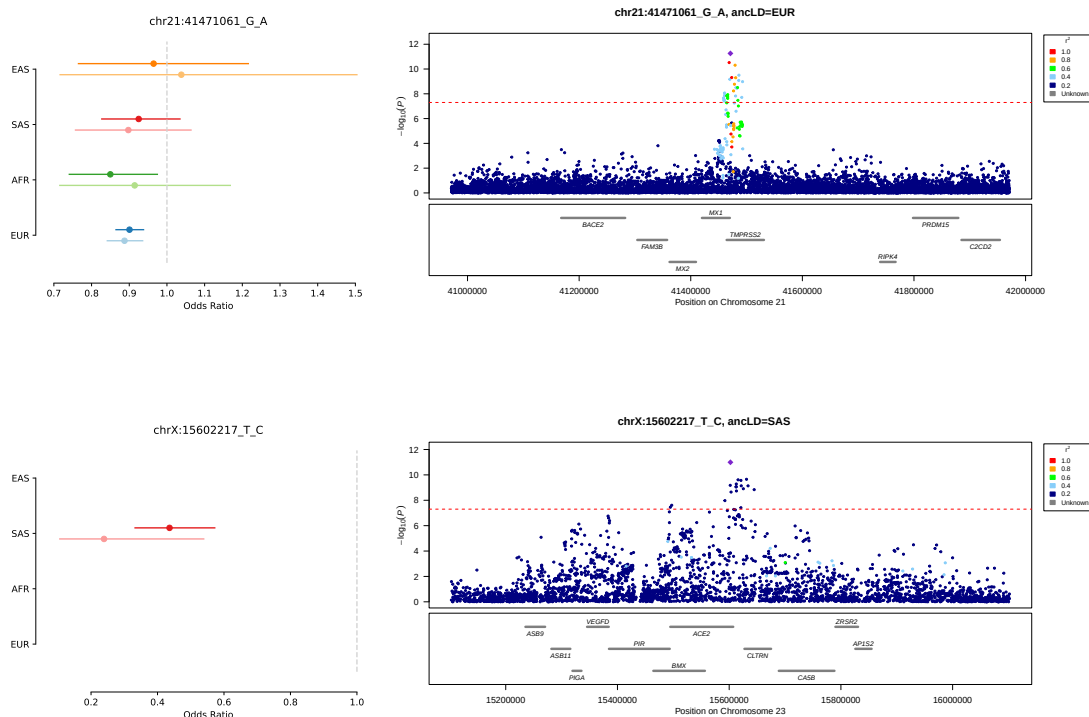


Figure 9: LocusZoom and forest plots illustrating all lead genetic association signals listed in Table 2. Panels on the right show LocusZoom plots centered on lead variant. P -values are from the cross-ancestry GenOMICC severe versus ctrl-all meta-analysis. The population used to calculate LD is indicated on the title and corresponds to the population having the smallest P -value in the GenOMICC meta-analysis. Panels on the left compare cross-ancestry odds ratios of lead variants from analyses using all controls (ctrl-all, darker lines) vs. analyses using only controls with mild COVID-19 (ctrl-mld, lighter lines).

3 Acknowledgments

Genomics England: This research was made possible through access to data in the National Genomic Research Library¹⁷, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library (<https://www.genomicsengland.co.uk/research>) holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. REACT: National Institute for Health and Care Research (NIHR) and UK Research and Innovation (UKRI) - REACT-Genomics England (REACT-GE) (MR/V030841/1) and REACT-Long COVID (REACT-LC) (COV-LT-0040). The REACT study was funded by the UK Department of Health and Social Care with supplemental funding from the Huo Family Foundation.

4 Conflicts of Interest

Erola Pairo-Castineira is an employee and shareholder of Regeneron Pharmaceuticals, Inc. No Regeneron resources were used in this publication.

5 Extended References

- [1] Pairo-Castineira, E. *et al.* GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature* **617**, 1–15 (2023).
- [2] Kanai, M. *et al.* A second update on mapping the human genetic architecture of covid-19. *Nature* **621**, 621:7977 **621**, E7–E26 (2023). URL <https://www.nature.com/articles/s41586-023-06355-3>.
- [3] Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human gwas trait-associated loci. *Nature Genetics* **53**, 1527–1533 (2021). URL <https://www.nature.com/articles/s41588-021-00945-5>.
- [4] Kanai, M. *et al.* Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell genomics* **2** (2022).
- [5] Kousathanas, A. *et al.* Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature* **607**, 97–103 (2022).
- [6] Ellinghaus, D. *et al.* Genomewide association study of severe covid-19 with respiratory failure. *The New England journal of medicine* **383**, 1522–1534 (2020).
- [7] The COVID-19 Host Genetics Initiative & Ganna, A. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. *medRxiv : the preprint server for health sciences* 2021.03.10.21252820 (2021).
- [8] Pathak, G. A. *et al.* A first update on mapping the human genetic architecture of covid-19. *Nature* **608**, E1–E10 (2022). URL <https://www.nature.com/articles/s41586-022-04826-7>.
- [9] Degenhardt, F. *et al.* Detailed stratified gwas analysis for severe covid-19 in four european populations. *Human Molecular Genetics* **31**, 3945–3966 (2022). URL <https://dx.doi.org/10.1093/hmg/ddac158>.
- [10] Gao, B. & Zhou, X. Mesusie enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nature Genetics* **56**, 170–179 (2024).

- [11] Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society. Series B, Statistical methodology* **82**, 1273–1300 (2020).
- [12] Bernabeu, E. *et al.* Sex differences in genetic architecture in the uk biobank. *Nature Genetics* 2021 53:9 **53**, 1283–1289 (2021). URL <https://www.nature.com/articles/s41588-021-00912-0>.
- [13] Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* 2021 53:7 **53**, 1097–1103 (2021).
- [14] Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: Study designs and statistical tests. *American Journal of Human Genetics* **95**, 5 (2014). URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC4085641/>.
- [15] Lee, S., Wu, M. C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics (Oxford, England)* **13**, 762–775 (2012). URL <https://pubmed.ncbi.nlm.nih.gov/22699862/>.
- [16] Liu, Y. *et al.* Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *American journal of human genetics* **104**, 410–421 (2019). URL <https://pubmed.ncbi.nlm.nih.gov/30849328/>.
- [17] National Genomic Research Library (2017).