

**Supplementary Table 1. Primary death cause in UKBiobank for people tested positive for COVID-19**

Primary cause of death#		Cause of death relevant to COVID-19*	Cause of death not relevant to COVID-19
ID	Description		
	Bronchus or lung, unspecified	2	
	Malignant neoplasm of ovary	1	
	Malignant neoplasm of prostate	1	
	Non-insulin-dependent diabetes mellitus- , With multiple complications	1	
	Unspecified dementia		1
	Alzheimer's disease, unspecified		1
	Chronic ischaemic heart disease, unspecified		1
	Bronchopneumonia, unspecified	1	
	COVID-19 virus identified	183	
	Other reconstructive surgery	1	
<b>Totals</b>		<b>190</b>	<b>3</b>

#UK Biobank classification for cause of death see [link](#)

\*Including people whose primary or non-primary cause of death is U071, and when cause of death is relevant to symptoms of COVID-19.

**Supplementary Table 2. Gender distribution and underlying health conditions of case and control sets**

<b>Characteristic</b>	<b>Case n=180 (Dead)</b>	<b>Case n=1141 (Not dead)</b>	<b>Approx. risk fold increase</b>
<b>Gender Male</b>	127	576	1.4
<b>Hypertension</b>	91	372	1.6
<b>Diabetes</b>	31	97	2
<b>High cholesterol</b>	49	156	2

**Supplementary Table 3. All missense variants included in the genes in UKBiobank Axiom Array**

	<b>Case (dead from virus, 180)</b>	<b>Control (1141)</b>
<b>rs150892504 ERAP2</b>	6	5
	M: 3, F: 3	M:3, F: 2
	2 hypertension and high cholesterol, 1 MS/chronic sinusitis, 1 with depression/anxiety 2 cancers (Malignant neoplasm of skin and Follicular non-Hodgkin's lymphoma)	1 no condition, 3 hypertension, 1 with irritable bowel
<b>rs138763430 BRF2</b>	4	2
	M: 3, F: 1	M:2
	1 Cancer (malignant neoplasm of skin) 2 Hypertension 1 No condition	1 hypertension, high cholesterol 1 asthma
<b>rs117665206 TMEM181</b>	5	6
	M: 5	M:2, F:4
	2 Cancer (Skin of trunk, Transverse Colon) All have hypertension	1 hypertension others: gastro-oesophageal, ankylosing spondylitis, asthma, appendicitis,back pain,vaginal prolapse
<b>rs147149459 ALOXE3</b>	4	3
	M:3, F: 1	M:1, F:2
	1 Cancer (Malignant neoplasm of thyroid gland) 2 hypertension 1 asthma/bipolar disorder	3 hypertension (including 1 Breast cancer)
<b>rs151256885 ALOXE3 intron</b>	3	0
	M:0	
	All hypertension, including 2 diabetes	

**Supplementary Table 4. P-values (in\_var\_meta\_p) in other GWAS studies**

Phenotype			ANA_A2_V2	ANA_B1_V2	ANA_B2_V2	ANA_C1_V2	ANA_C2_V2	ANA_D1_V2
			very severe respiratory confirmed covid vs. population	hospitalized covid vs. not hospitalized covid	hospitalized covid vs. population	covid vs. lab/self-reported negative	covid vs. population	predicted covid from self-reported symptoms vs. predicted or self-reported non-covid
<b>case</b>			536	928	3199	3523	6696	1865
<b>Control</b>			329391	2028	897488	36634	1073072	29174
<b>Population**</b>			EUR	EUR	EUR	MIX	MIX	EUR
<b>rs150892504</b>	T/C	ERAP2	NA*	1.456E-01	2.12E-02	4.984E-01	8.195E-01	5.762E-02
<b>rs138763430</b>	T/C	BRF2	NA	4.841E-01	6.161E-01	9.507E-01	7.782E-01	4.009E-02
<b>rs117665206</b>	T/C	TMEM181	NA	7.373E-01	2.246E-01	1.692E-02	8.203E-02	6.682E-01
<b>rs147149459</b>	A/G	ALOXE3	NA	NA	7.823E-01	7.278E-01	5.922E-01	4.143E-01
<b>rs151256885</b>	T/C	ALOXE3 intron	3.69E-01	NA	1.466E-01	4.44E-01	6.761E-01	NA

Data obtained from: <https://www.covid19hg.org/results/>

\*NA. Variant not found. \*\*Majority of population used; more information is available from the above link.

**Supplementary table 5. Detailed ethnic breakdown by ethnicity of the variants in various population from different studies.**

SNPs	Study	% of population by ethnic group															
		All		Ashkenazi Jewish		American		East Asian		African		European		Mexican		South asian	
		n	Alt Allele	n	Alt Allele	n	Alt Allele	n	Alt Allele	n	Alt Allele	n	Alt Allele	n	Alt Allele	n	Alt Allele
rs150892504 C>T	gnomAD	31398	0.0022	290	0.041	848	0	1560	0	8714	0.0002	18900	0.00265	NA	NA	NA	NA
	PAGE	78700	0.00133	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10810	0.00315	856	0.016
	ExAC	121408	0.005214	NA	NA	11578	0.00207	NA	NA	10406	0.00029	73352	0.00449	NA	NA	NA	NA
rs138763430 C>T	gnomAD	31404	0.00248	290	0	848	0.002	1560	0	8718	0.0005	18900	0.00365	NA	NA	NA	NA
	PAGE	78696	0.0017	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10810	0.00222	856	0
	ExAC	98570	0.00267	NA	NA	9130	0.0023	NA	NA	750	0.004	58788	0.004	NA	NA	NA	NA
rs117665206 C>T	gnomAD	31390	0.00682	290	0.007	848	0.002	1558	0	8710	0.0024	18896	0.00921	NA	NA	NA	NA
	PAGE	78696	0.00241	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10810	0.00296	856	0.002
	ExAC	120752	0.006087	NA	NA	11576	0.00199	NA	NA	9800	0.0011	73340	0.00885	NA	NA	NA	NA
rs147149459 G>A	gnomAD	31388	0.00124	290	0	848	0	1556	0	8706	0.0001	18900	0.00196	NA	NA	NA	NA
	PAGE	78694	0.00027	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	10810	0	856	0
	ExAC	121404	0.000939	NA	NA	11576	0.00035	NA	NA	10406	0.00029	73348	0.00145	NA	NA	NA	NA
rs151256885 C>T	ALFA	17562	0.00512	NA	NA	NA	NA	NA	NA	676	0.033	14270	0.00301	NA	NA	4	0

Data obtained from: <https://www.ncbi.nlm.nih.gov/snp/>

NA: Data not available for this variant in this population under this study

# Non-coding Intron Variant Prediction

## regSNP-intron

The full result with all the features can be downloaded here: [snp.prediction.txt](http://snp.prediction.txt).



Output from the regSNP-intron software for snp rs151256885 which lies in the intronic region of ALOXE3.

## Structural Analysis Methods

### ERAP2

ERAP2 gene information was retrieved from UniProtKB entry Q6P179. For the purpose of structural analysis of the mutation R751C found in ERAP2 gene, we used the structure of endoplasmic reticulum aminopeptidase solved at 2.50 Å resolution (PDB ID 5AB0 obtained from the UniProtKB entry). The wildtype residue R751 was found to be solvent inaccessible and buried at the residue depth of 5.3 Å calculated using the program DEPTH<sup>53</sup>. The buried environment was also found to be well packed with the occluded surface packing (OSP)<sup>54</sup> density of 0.5. The side chain atoms in the wildtype residue R751 engage in a total of 17 direct polar and non-polar contacts with the neighbouring atoms within 10 Å radius (Figure 1A - top right). The point mutation significantly reduces the number of contacts to 5 suggesting a destabilising effect on the protein structure (Figure 1A – bottom right).

### ALOXE3

The protein structural data was not available for the ALOXE3 gene. So, in order to understand the impact of mutation we used a homology model derived from the template PDB structure (PDBID 4NRE) at 55% sequence identity. Template search and homology model building was done using SWISS-MODEL<sup>55</sup>. Starting from the homology model, the model of the mutant was built and optimised using the MODELLER package<sup>56</sup>. The wildtype residue P630 was completely buried at the residue depth of 11.3 Å and has an OSP value of 0.53. A total of 10 contacts were observed (Figure 1A – top right). Interesting, mutant P630L represents a bulky replacement in an already well packed environment. Due to the substitution of larger amino acid Leucine, the number of contacts has gone up to 22 in the mutant structure at the expense of two serious steric clashes with the neighbouring residues S608 and I634 (shown as red lines in Figure 1B – bottom right). Bulky mutations in a buried and well packed environment are usually not tolerated and lead to protein destabilisation.

### TMEM181

No structure was found for the TMEM181 gene representing a transmembrane protein. Homology model building was challenging due to the lack of good template structures with at least more than 30% sequence identity. In fact, for all of the top templates identified by SWISS-MODEL, the sequence identity was less than 16%. Sequence based prediction using TMHMM<sup>57</sup> analysis showed the presence of 9 transmembrane helices. The mutant R403C is found on the 5<sup>th</sup> transmembrane helix close to the intra-cellular membrane region (Figure 1C). Sequence analysis showed that at least 74% all transmembrane helix residues predicted by TMHMM were hydrophobic. Polar and charged residues like arginine were also found in the membrane helix predominantly located close to the end of the helix that are proximal to the intra or extra cellular region. Due to the lack of structural data, it is difficult to predict the orientation of the R403 with respect to the membrane pore and explain the structural impact of R403C mutation. However, buried and charged arginine residues have been suggested to form salt bridges with lipid bilayer (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2173043/>). In general, buried charged residues are involved in catalysis, ligand binding or in stabilising helix caps found at or close to the ends of the helices (<https://pubmed.ncbi.nlm.nih.gov/11080642/>) which is consistent with the location of R403C mutant.

## **BRF2**

Homology model was built for the Zn Ribbon domain of BRF2 gene (residues 3-40) using SWISS-MODEL. NMR structure of the N-terminal domain of TFIIB (PDB ID 1PFT) with 36% sequence identity was used as the template to build the model. The mutant D9N was found proximal to the zinc binding site coordinated by 4 well conserved cysteine residues (Figure 1C – left). Upon mutation, no significant structural deviation has been observed between the wildtype and mutant structure. However, comparison of the overall electrostatic potential surface area shows a visible difference in the charge distribution with the mutant being less negatively charged (Figure 1C – right) compared to the wildtype (Figure 1C – center). This change would impact the ability of the Zn Ribbon domain to recognise its nucleic acid binding partners.

## **Ethnic breakdown**

ONS data here (not sure you can copy the graph for publication since copyright). But present some numbers at least for the major categories (adjusted for socio – demographic issues)

Table of more detailed breakdown by ethnicity of our variants in populations from 1000G including pakistani/bangladeshi and others. And/or get finer detail from e.g. <https://genomeasia100k.org/> and dbSNP