medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Ethnically diverse mutations in *PIEZO1* associate with SARS-CoV-2 positivity

Cheng, C.W., Deivasikamani, V., Ludlow, M.J., De Vecchis, D., Kalli, A.C., Beech, D.J., Sukumar, P.

School of Medicine, University of Leeds, Leeds, LS2 9JT, UK.

*Author for correspondence: Professor David J Beech, Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, LIGHT Laboratories, Clarendon Way, University of Leeds, Leeds, LS2 9JT, United Kingdom. Telephone +44 (0) 113 3434323. Email: D.J.Beech@leeds.ac.uk.

ABSTRACT

COVID-19, caused by the SARS-CoV-2 virus, carries significant risk of mortality and has spread globally with devastating societal consequences. Endothelial infection has been identified as a feature of the disease and so there is motivation to determine the relevance of endothelial membrane mechanisms affecting viral entry and response. Here, through a study of patient data in UK Biobank released on 16 April 2020, we suggest relevance of PIEZO1, a non-selective cation channel protein that both mediates endothelial responses to mechanical force and unusually indents the cell membrane. PIEZO1 notably has roles that may also be relevant in red blood cell function, pulmonary inflammation, bacterial infection and fibrotic auto-inflammation. We provide evidence that single nucleotide polymorphisms (SNPs) in the gene encoding PIEZO1 are more common in individuals who test positive for SARS-CoV-2 regardless of pre-existing hypertension, myocardial infarction, stroke, diabetes mellitus or arthritis. Some of these SNPs are more common in African and Caribbean populations, which are groups that were recently shown to have greater susceptibility to infection. One of the SNPs is a missense mutation that results in an amino acid change in an evolutionarily conserved and previously unexplored N-terminal region PIEZO1. The data support the notion of genetic factors influencing SARS-CoV-2 infection and suggest a specific role for PIEZO1.

INTRODUCTION

Coronavirus disease 19 (COVID-19) is caused by a new virus called SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2)[1]. In 2020 this virus triggered a global health crisis, spreading rapidly with no vaccine or treatment yet available[2,3]. One way to address the problem could be to understand how the virus enters host cells because such knowledge could usefully inform repurposing of existing therapeutics and enable development of new therapeutics that reduce the dangers of the virus. There is already evidence that surface membrane proteins such as ACE2 and TMPRSS2 are important in SARS-CoV-2 entry[4] but there is little information on the roles of other membrane proteins, membrane lipid components or membrane structure. Research on other viruses has suggested that mechanisms of this type affect viral entry[5,6]. Particular membrane proteins of interest are ion channels[7,8], which embed in the membrane and enable flux of ions such as Ca^{2+}, a key intracellular signal[9] and regulator of coronavirus mechanisms[10-12].

A recent report on blood vessels of COVID-19 patients pointed to importance of endothelial cell infection and endotheliitis[13]. The study included investigation of the mesenteric vasculature because of circulatory collapse in which there was mesenteric ischaemia requiring surgical resection to remove part of the small intestine[13]. Viral elements were found in endothelial cells and there were accumulations of inflammatory cells alongside endothelial and inflammatory cell death[13]. The authors suggested that therapeutic strategies aimed at stabilising the endothelium may be useful, especially in patients who are vulnerable to COVID-

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

19 because of pre-existing endothelial dysfunction, including males and smokers and people who are obese or have hypertension or diabetes[13]. An independent study of lungs from patients who died from COVID-19 found severe endothelial injury and angiogenesis, contrasting with the lungs of patients who died from acute respiratory distress syndrome secondary to influenza[14]. In support of these conclusions, SARS-CoV-2 has been found to directly infect human blood vessel organoids[15]. Therefore the idea exists that vascular endothelium is particularly vulnerable to SARS-CoV-2 and important in COVID-19 severity.

For these reasons it could be helpful to learn specifically about the membrane mechanisms of endothelial cells that confer SARS-CoV-2 susceptibility and determine downstream consequences of the virus in the vasculature. In 2014, an intriguing ion channel protein called PIEZO1 was reported to be important in endothelium[16,17]. This protein forms Ca^{2+}-permeable non-selective cation channels that have extraordinary capability to respond to membrane tension[18] and shear stress caused by fluid flow along the endothelial membrane surface[16]. Unusually for membrane proteins, PIEZO1 indents the membrane in an inverted dome-like fashion and therefore modifies the overall structure of the membrane[19,20]. The channel shows marked activity in mesenteric endothelium[21] and there is increasing evidence of its importance in many aspects of endothelial function, such as angiogenesis[16,22] and pulmonary vascular permeability[23,24]. There are also roles in cardiovascular health and disease more generally[25], including in the regulation of interleukin-6[26], which is a key inflammatory mediator of COVID-19[27]. We know there is importance in humans because naturally occurring loss-of-function mutations in *PIEZO1* are known to be associated with lymphatic endothelial dysfunction[28,29] and varicose veins[30] and gain-of-function mutations associate with anaemia and malarial protection due to importance of *PIEZO1* in red blood cells[31-33]. Preliminary genetic analysis has hinted at numerous additional conditions that may be related to or exacerbated by mutation in *PIEZO1*[25].

Therefore we hypothesised that mutations in *PIEZO1* might relate to COVID-19 and so we tested this hypothesis using newly available data in the UK Biobank. UK Biobank is a health research resource that recruited over 500,000 people aged 40 to 69 years between 2006 and 2010 across the UK[34]. It recently incorporated SARS-CoV-2 infection data.

RESULTS

On 16 April 2020 the UK Biobank released the data of 1409 individuals tested for SARS-CoV-2 across its 22 assessment centres. A candidate gene association study was conducted to establish the link between SARS-CoV-2 infection (COVID-19) and *PIEZO1* gene. Table 1 shows the demographic properties of the 1409 study participants who were tested. Of those, 636 (45%) tested positive. The mean age of all individuals was 69.5 years and 756 individuals (54%) were men while 653 (46%) were women. 48% and 42% of men and women tested were found to be positive respectively. In the downstream analysis, we focused on the individuals who self-identified as “British” (82%). We evaluated the co-morbidities of these individuals obtained from non-cancer illness code (ID 20002). Of the total study sample of 1409 individuals, 521 (37%) self-reported that they had a diagnosis of hypertension, 77 (5%) myocardial infarction, 39 (3%) stroke, 123 (9%) diabetes mellitus and 214 (15%) arthritis.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Characteristic	*N*	COVID-19 Positive	COVID-19 Negative
Number of Individuals	1409	636	773
Age (Mean ± SD)	69.52 ± 8.74	69.31 ± 8.80	69.68 ± 8.74
Gender			
Male	756	363	393
Female	653	273	380
Ethnicity			
British	1158	494	664
others	251	142	109
Co-Morbidity			
Hypertension	521	251	270
Myocardial Infarction	77	33	44
Stroke	39	21	18
Diabetes mellitus	123	59	64
Arthritis	214	87	127

Table 1. Descriptive statistics and demographic data for COVID-19 in UK Biobank. *N* = Total number of individuals tested for COVID-19 in UK Biobank; SD = standard deviation. British ethnicity refers to the individuals who identified or reported themselves as "British". Others represent individuals who did not report as "British".

Two general models of logistic regression analysis assuming an additive genetic approach were conducted without adjusting for self-reported illnesses. The first general model was conducted on all individuals released by UK Biobank (*N* = 1409). To account for genetic heterogeneity amongst different ethnic populations in UK Biobank, the second logistic analysis was restricted to individuals who self-reported as "British". Both analyses were adjusted for age, sex, duration of moderate physical activity and the first ten principal components. Associations with COVID-19 risks were quantified using odds ratios (ORs) derived from logistic regression. Table 2 shows the synonymous and missense variants that were detected from the general model analysis.

The first general model included all the individuals and was adjusted for 13 covariates. Twenty variants were found to be statistically significant (*P*-value <0.05), indicating a degree of association. Of those, one was missense (rs1803328, OR = 0.805 [0.663 - 0.978]), three were synonymous (rs8057031, OR = 0.724 [0.602 – 0.871]; rs2290902, OR = 0.729 [0.539 - 0.985]; rs6500491, OR= 0.732 [0.540 - 0.992]) and 16 were intronic (Table 2 and Supplementary file).

To reduce the influence of different genetic backgrounds in the association analysis, the second general model was restricted to self-reported British among the 1409 individuals. In total, 1158 individuals were eligible for the next step of analysis using logistic regression and adjusted for 13 covariates. After adjusting for the covariates, 35 variants achieved the *P*-value threshold <0.05, indicating a significant association. One missense variant (rs7184427, OR = 0.761 [0.581 - 0.996]), six synonymous variants (rs8057031, OR = 0.757 [0.625 - 0.917]; rs2290902, OR = 0.627 [0.454 - 0.867]; rs6500491, OR = 0.622 [0.448 - 0.863]; rs1061228, OR = 1.403 [1.060 - 1.85]; rs8043924, OR = 0.720 [0.532 - 0.974]; rs4782430, OR = 0.718 [0.528 - 0.978]) and 28 intronic variants (Table 2 and Supplementary file). The missense variant rs1803382 was significant when the association analysis included the entire population (*n* = 1409) but became non-significant (*P*-value = 0.254) when the analysis was confined to

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

British. In contrast, three synonymous variants became statistically significant when the analysis was confined to British population (rs1061228, rs8043924, rs4782430).

SNP ID	BP	REF	ALT	A1	Mutation	General Model 1[a]			General Model 2[b]		
						P-value	OR	95% CI	*P*-value	OR	95% CI
rs8057031	88783449	C	G	C	Synonymous	**<0.05**	0.724	0.602 – 0.871	**0.004**	0.757	0.625 - 0.917
rs1803382	88783100	T	C	C	Missense	**0.029**	0.805	0.663 - 0.978	0.254	0.892	0.732 - 1.086
rs2290902	88782676	A	G	A	Synonymous	**0.040**	0.729	0.539 - 0.985	**0.005**	0.627	0.454 - 0.867
rs6500491	88783521	T	C	T	Synonymous	**0.044**	0.732	0.540 - 0.992	**0.004**	0.622	0.448 - 0.863
rs1061228	88782079	G	A	A	Synonymous	0.053	1.308	0.997 -1.718	**0.018**	1.403	1.060 - 1.85
rs8043924	88787704	G	C	G	Synonymous	0.228	0.841	0.635 - 1.114	**0.033**	0.720	0.532 - 0.974
rs4782430	88792047	A	G	A	Synonymous	0.321	0.864	0.648 - 1.153	**0.035**	0.718	0.528 - 0.978
rs7184427	88804734	A	G	A	Missense	0.082	0.798	0.619 -1.029	**0.047**	0.761	0.581 - 0.996

Table 2. *PIEZO1* mutations (single nucleotide polymorphisms, SNPs) associated with COVID-19 in the General Model. BP = base pair; REF = reference allele; ALT = alternate allele; A1 = tested allele; OR = odds ratio; CI = confidence intervals.
[a] Logistic regression model with 1409 individuals tested for COVID-19 in UK Biobank. Adjusted for age, sex, duration of moderate activity, principal component 1 to 10.
[b] Logistic regression model of 1158 individuals who self-reported as "British". Adjusted for age, sex, duration of moderate activity, principal component 1 to 10.

To extend our analysis of the "British" cohort, the association tests were adjusted to five underlying types of disorder; hypertension, myocardial infarction, stroke, diabetes mellitus and arthritis. The purpose of adjusting the analyses to these disorders was to account for covariate effects, to test if the SNPs reflected true associations between *PIEZO1* and COVID-19 status. The COVID-19 case status was the response variable while the imputed genotype was the predictor variable using PLINK 2.0 logistic regression model. All the models were adjusted to age, sex, duration of moderate activity, along with the first ten principal components to control population stratification. Across the five models of logistic regression in Table 3, the results revealed similar missense and synonymous variants that achieved statistical significance (*P*-value <0.05). Additionally, the analysis provided evidence supporting the missense variant (rs7184427), which we observed in General Model 2. The missense variant was consistently significant across the five models with mean odds ratio of 0.76. Besides, six synonymous variants persisted across the analyses after adjusting for the covariates (rs8057031, mean OR = 0.7276; rs6500491, mean OR = 0.621; rs2290902, mean OR = 0.653; rs1061228, mean OR = 1.403; rs8043924, mean OR = 0.718; rs4782430, mean OR = 0.76). The odds ratio and *P*-value of each variant across the five models did not deviate significantly.

SNP ID	REF	ALT	A1	Mutation	Model 1[a]		Model 2[b]		Model 3[c]		Model 4[d]		Model 5[e]	
					P-value	OR	P-value	OR	P-value	OR	P-value	OR	P-value	OR
rs8057031	C	G	C	Synonymous	0.004	0.756	0.004	0.756	0.004	0.756	0.004	0.757	0.004	0.613
rs6500491	T	C	T	Synonymous	0.004	0.621	0.004	0.622	0.004	0.622	0.005	0.624	0.004	0.618
rs2290902	A	G	A	Synonymous	0.005	0.626	0.005	0.627	0.005	0.627	0.005	0.629	0.005	0.758
rs1061228	G	A	A	Synonymous	0.018	1.403	0.018	1.403	0.018	1.403	0.017	1.409	0.020	1.396
rs8043924	G	C	G	Synonymous	0.033	0.719	0.033	0.720	0.033	0.720	0.035	0.722	0.027	0.710
rs4782430	A	G	A	Synonymous	0.034	0.717	0.035	0.718	0.035	0.718	0.037	0.721	0.029	0.709
rs7184427	A	G	A	Missense	0.046	0.760	0.047	0.761	0.047	0.761	0.049	0.762	0.040	0.754

Table 3. *PIEZO1* mutations (single nucleotide polymorphisms, SNPs) significantly associated with COVID-19 after adjusting the self-reported illnesses. REF = reference allele; ALT = alternate allele; A1 = tested allele; OR = odds ratio; all models data were adjusted for age, sex, duration of moderate activity, principal component 1 to 10 in addition to the specific disease condition as given below.
[a] Adjusted for hypertension status.
[b] Adjusted for myocardial infarction status.
[c] Adjusted for stroke status.
[d] Adjusted for diabetes mellitus status.
[e] Adjusted for arthritis status.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The relationships of 16 PIEZO1 orthologous protein sequences were inferred by constructing a maximum likelihood tree using RAxML (Figure 1A). Amino acid sequence alignment of PIEZO1s revealed a high degree of conservation from residue 243 to 258 (Figure 1B). The missense variant, rs7184427, encodes an amino acid located at position 250 in human PIEZO1 (highlighted using the brown box in Figure 1B). This residue is completely conserved across diverse species. It is important to note that the analysis used the reference genome, which does not always employ the major allele, as in this case for humans. Our data show that in many humans, alanine (A) is encoded by the major allele and is therefore common at this site.

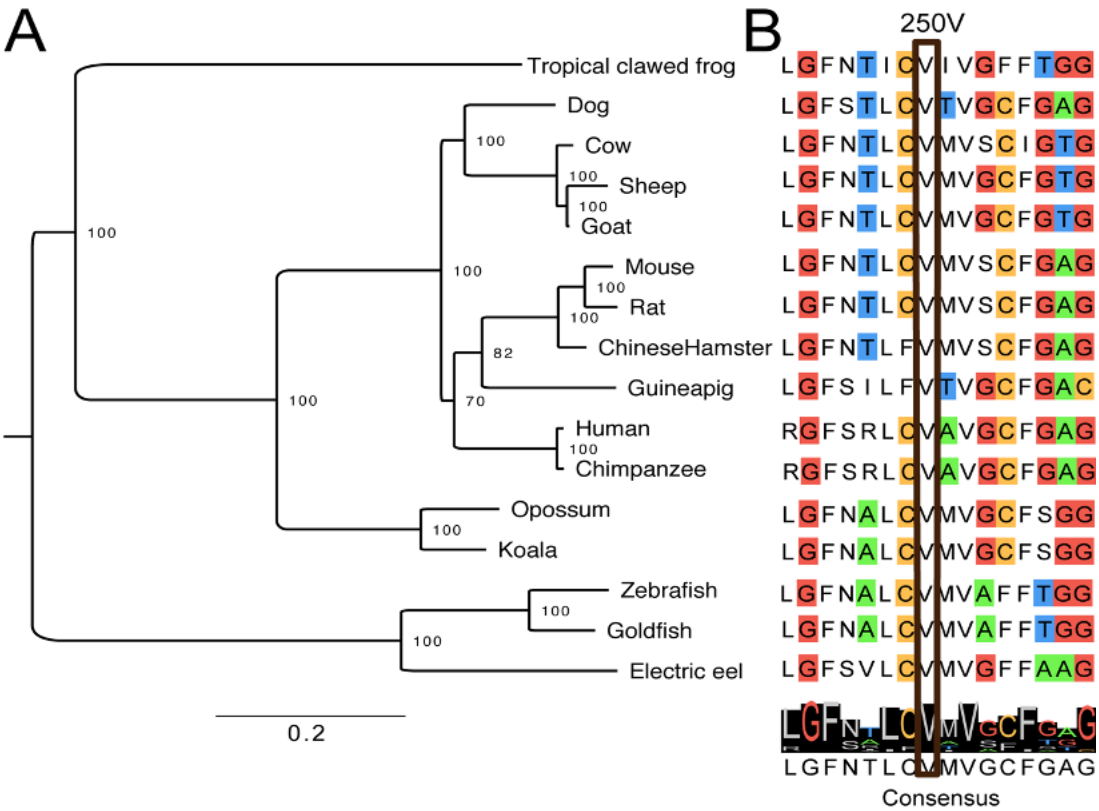

Figure 1. Phylogenetic tree of PIEZO1 sequences across 16 species. (**A**) The phylogenetic tree was inferred from maximum likelihood in RAxML using 100 bootstrap replicates. The tree was midpoint rooted with support values showing on the branch. All nodes achieved >70% bootstrap support. Scale bar represents 0.2 substitutions per site. (**B**) Sequence alignment for amino acids (single-letter code) at the V250 position. See the text regarding A250. The consensus amino acid is shown at the bottom.

Following the logistic regression analyses, we focused on the missense variant (rs7184427) that remained significant in all analyses. The major (G) allele was associated with greater susceptibility to COVID-19. The variant was located in the N-terminal region of the PIEZO1 protein (Figure 2). Although there are 3D structural data for the mouse PIEZO1[19,35], none of these structures includes the N-terminal region of PIEZO1 in which A or V at position 250 is located. For this reason, using the available structural data we modelled the first three 4-helical bundles of PIEZO1 in one of the experimentally derived 3D structures of PIEZO1. Our model enables us to suggest that V250 (V257 in mouse PIEZO1) is located near the N-terminal region of the blades, at the cytosolic side of the protein.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

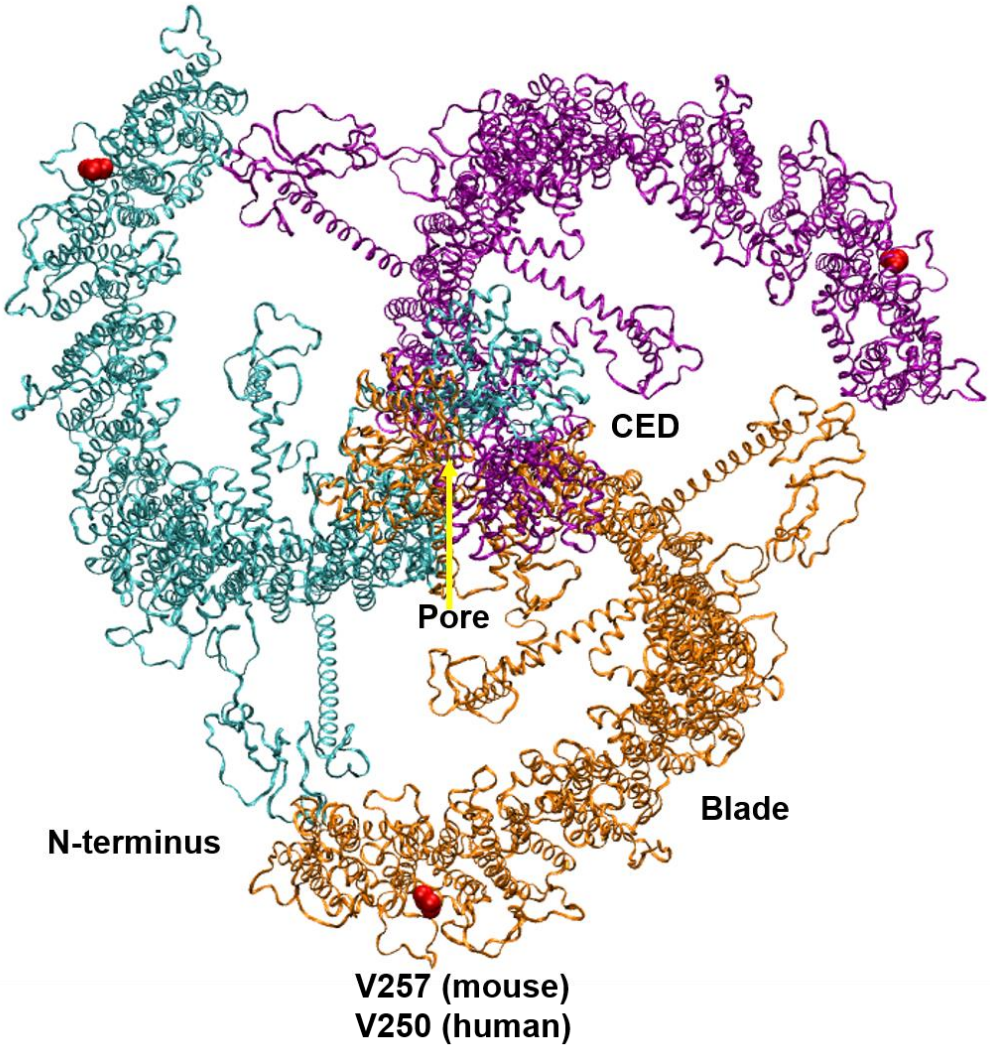

Figure 2. Molecular model of the PIEZO1 channel indicating the position of the amino acid residue affected by rs7184427. The model is for mouse PIEZO1 because structural data exist for this PIEZO. Features of the channel are labelled: the C-terminal extracellular domain (CED), the ion permeability pore (Pore) and one of the 3 propeller blades (Blade). The three identical PIEZO1 subunits required to make a single PIEZO1 channel are shown in purple, orange and cyan. Valine 250 (V257 in the mouse PIEZO1) is shown in red. See Chong et al[20] for details of how this model was constructed.

The frequency of the major allele (the pro-COVID-19 G allele resulting in V250A mutation) was 84.4%, 82.0% and 82.9% in global populations of UK Biobank, 1000 Genome and Genome Aggregation database (gnomAD) (Table 4). The frequency in the South Asian populations was lower than in the global population (65.0-70.8%) but higher in the African and East Asian populations (87.1-90.1%). Pro-disease allele frequencies of all of the identified synonymous mutations also varied with ethnicity (Figure 3). In two of these cases (rs8057031 and rs4782430) there was again higher frequency in African and Chinese populations (Figure 3).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Population	Major Allele Frequency
UK Biobank	
All populations (*n* = 487,410)	0.844
Males (*n* = 222,995)	0.844
Females (*n* = 264,303)	0.844
African (*n* = 3,211)	0.845
Bangladeshi (*n* = 223)	0.705
British (*n* = 430,630)	0.846
Caribbean (*n* = 4,303)	0.879
Chinese (*n* = 1,506)	0.878
Indian (*n* = 5,665)	0.696
Pakistani (*n* = 1,750)	0.708
1000 Genome Database	
Global (*n* = 3,500)	0.820
African (*n* = 1,018)	0.871
American (*n* = 535)	0.823
East Asian (*n* = 617)	0.901
European (*n* = 669)	0.837
South Asian (*n* = 661)	0.650

Population	Major Allele Frequency
Genome Aggregation Database (gnomAD)	
All populations (*n* = 71,702)	0.829
African (*n* = 21,042)	0.884
American (*n* = 6,835)	0.839
Ashkenazi Jewish (*n* = 1,662)	0.863
East Asian (*n* = 1,567)	0.880
Finnish (*n* = 5,244)	0.852
Non-Finnish European (*n* = 32,299)	0.853
Others (*n* = 1,077)	0.851
South Asian (*n* = 1,526)	0.693

Table 4. Major allele frequency (pro-COVID-19 G allele) of rs7184427 across different populations.

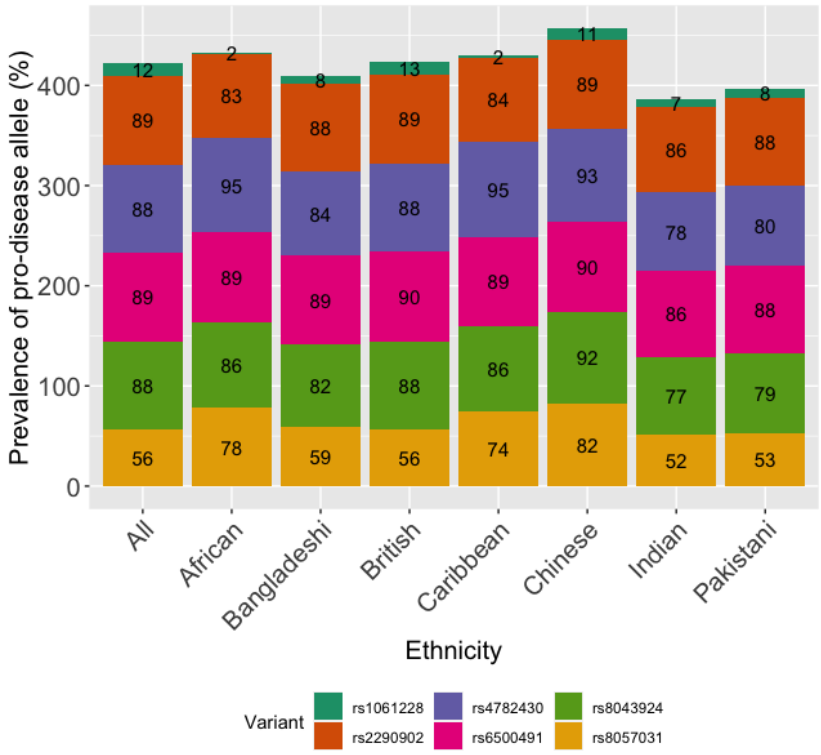

Figure 3. Pro-COVID-19 allele frequency (% total) of synonymous mutations across different populations in UK Biobank.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

DISCUSSION

Our data suggest genetic variation in *PIEZO1* that increases susceptibility to SARS-CoV-2 infection. Such variation could increase the risk of other diseases, thereby indirectly affecting SARS-CoV-2 infection, but our analysis addressed major relevant diseases and pointed instead to PIEZO1 as a direct determinant of infection rate. In total, 1 nonsynonymous missense SNP and 6 synonymous SNPs were found to be linked to SARS-CoV-2 infection rates. To the best of our knowledge these data are the first to provide evidence of genetic linkage to SARS-CoV-2 infection and the first to implicate PIEZO1 as a factor in this disease.

The prevalence of many of these SNPs varies with ethnicity and this could be a factor underlying diversity seen in SARS-CoV-2 infection rates in different populations. The missense SNP (rs7184427) associated with more infections in the White British group is the G allele. It is the major allele and most common generally in the White British population and human populations in general (82-84% of people). It encodes alanine at position 250 in PIEZO1 protein, whereas the remaining people with the minor allele have valine at this position. The major allele is more frequent in African and Chinese (East Asian) groups and less frequent in South Asian groups. Some of the synonymous SNPs also show such variability, with rs8057031 and rs4782430 being more frequent in African / Caribbean and Chinese groups. Again lower frequency is sometimes seen in South Asian groups.

Data are emerging on the relevance of ethnicity to SARS-CoV-2 infection rates and the main conclusion so far is that infection is more common in Black compared with White people, regardless of socioeconomic conditions and other confounding factors[36,37]. This suggests greater genetic susceptibility of Black people. Therefore, the greater frequency of some of the pro-disease *PIEZO1* SNPs in African and Caribbean groups could be an explanation, or at least contributory factor. We are not aware of evidence that infection rates are different in Asians. South Asian groups in the UK have experienced greater lethality from COVID-19[38] but the factors determining this effect may be different from those determining infection rates.

We do not yet know if alanine in place of valine at position 250 in PIEZO1 causes less, more or no change in the properties of the PIEZO1 channels or whether the altered DNA sequence affects *PIEZO1* expression. It will take time and investment to determine such matters. The valine amino acid is similar to the alanine amino acid but the chemical difference may be crucial, for example in α-helical integrity[39]. Evolutionary preservation of this residue from fish to human (Figure 1) suggests importance, otherwise variations at this site would likely have occurred and been tolerated and therefore led to divergence over millions of years.

While we have insight into how some parts of the PIEZO1 molecular machine operate[19,35,40], the particular region of PIEZO1 incorporating V or A at position 250 is uncharted territory[20]. Progress with cryo-EM techniques has led to structural information[19,35] but the N-terminal region has not been resolved, possibly because it has highly flexible structural conformations. Nor has this region been studied by laboratory mutagenesis studies. Based on our molecular modelling we speculate that V250A is important in how the tip of the propeller blade curves round to abut the next blade (Figure 2). It may also be important in how the blades and the entire channel interact with membrane lipids and nearby proteins and therefore how the membrane structure is shaped and the sensitivity to mechanical perturbation is determined[20]. The molecular and functional studies required to unravel the roles of this blade tip and specifically the effect of V250A will be challenging. Moreover, our studies have shown the importance of understanding PIEZO1 in its native membrane environment; the native endothelial and red blood cell environments profoundly alter the gating characteristics of PIEZO1[16,25,41,42]. We do not yet know whether endothelial cell dysfunction affects the properties of PIEZO1, but any changes could be relevant to SARS-CoV-2.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

There is a public database entry indicating a link of V250A to hereditary lymphedema in one patient[43] and we know that generalised lymphatic dysplasia can be caused by loss of PIEZO1 expression[28,29]. Therefore we tentatively speculate that V250A reduces PIEZO1 function and that this is the reason for increased susceptibility to SARS-CoV-2 infection. If this is the case, PIEZO1 agonists may be a route to rescuing normal channel activity and protecting against infection. This may be a safe approach because we know that PIEZO1 gain-of-function is tolerated in people and even confers protective advantage in large populations where malaria is endemic[31-33]. A screen of 3.25 million small-molecules identified a PIEZO1 agonist called Yoda1[44]. It is increasingly used successfully as a tool compound in experimental studies of PIEZO1[25] and there is on-going work to elucidate the chemical structure-activity requirements and improve the physicochemical properties while retaining efficacy; some of these studies have been published[45]. To the best of our knowledge there is currently no PIEZO1 agonist that would be suitable for administration to people but the available data suggest that the principle of chemical enhancement of PIEZO1 is possible and thus that a therapeutic drug targeted to PIEZO1 is a realistic consideration.

In addition to a possible link to lymphedema, our study raises the question of whether COVID-19 is linked to varicose veins. A genome-wide association study identified mutations in *PIEZO1* as determinants of varicose veins[30] and a recent study specifically linked V250A to varicose veins[46]. These observations further encourage the idea of an important relationship between vascular integrity and COVID-19.

PIEZO1 may act in this context through impact on membrane structure, thus affecting viral entry, but we should also consider the property of PIEZO1 to confer Ca^{2+} permeability on the membrane because transmissibility of both SARS-CoV and MERS-CoV is Ca^{2+}-dependent and both are closely related to SARS-CoV-2[10,11,47]. E protein of SARS-Co-V forms Ca^{2+} permeable ion pores on ER/Golgi membrane and its activity can drive the inflammation and severe respiratory distress syndrome by enhancing IL-1β production[48]. One of the accessary proteins of SARS-CoV undergoes conformational changes upon binding to Ca^{2+} and significantly contributes to disease progression[12,49]. Future studies should address these possibilities.

Synonymous mutations do not change the amino acid sequence but they could still, in theory, have effects by altering expression of the *PIEZO1* gene, thereby affecting the total amount of PIEZO1 available for functional consequence. We note that one of the synonymous mutations we identified (rs2290902) has also been associated with HIV-1 infection[50] but we are not aware of further investigation of this matter. It is nevertheless consistent with our general idea that PIEZO1 is relevant to viral infection.

In addition to its endothelial and red blood cell functions, PIEZO1 is expressed in other cell types and implicated in other aspects of mammalian biology[25,40]. It may be particularly important to note in this context the recent persuasive evidence for a role of PIEZO1 in pulmonary inflammation, bacterial infection and fibrotic auto-inflammation[51]. In this study it was concluded that stimulation of PIEZO1 in immune cells by cyclical mechanical force is essential for innate immunity. Therefore we speculate that any loss of PIEZO1 function would increase susceptibility to lung infection and perhaps other infections and that PIEZO1 agonism might be beneficial. Cyclical activation of PIEZO1 might also be important in the protective benefits of physical exercise where PIEZO1 has been suggested to have an important role in blood pressure elevation of exercise that increases performance[21]. PIEZO1 might be relevant to the speculation about a connection between malarial protection and SARS-CoV-2 infection[52,53] because heterozygosity for PIEZO1 gain-of-function mutation has been suggested to be a major factor mediating protection from malaria[33].

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

We recognise that a potential limitation of this study is the relatively small size of the SARS-CoV-2 data set on which we could base the analysis at this time. Analysis of additional data, as they becomes available, will be important for deeper understanding and exploration of *PIEZO1* specifically in non-White ethnic groups. Reopening of research laboratories will enable experimental investigation of underlying mechanisms and the testing of therapeutic strategies.

In summary, this early study suggests genetic association that is relevant to COVID-19 and ethnic variation in infection rates and a previously unrecognised relationship between COVID-19 and *PIEZO1*, a gene that encodes an important mechanically-activated ion channel.

METHODS

Demographic information

The UK Biobank resource recruited over 500,000 people aged 40 to 69 years between 2006 and 2010 across the UK[34]. Participants completed a detailed clinical, demographic and lifestyle questionnaire, underwent clinical measures, provided biological samples (blood, urine and saliva) for future analysis and agreed to have their health records accessed. In July 2017, the genetic information from 501,708 samples was released to UK Biobank research collaborators. The UK Biobank has now released the first cohort of individuals tested for COVID-19 (N = 1474). These individuals were used in the present analysis to study the association between COVID-19 and *PIEZO1*. Baseline assessments of these individuals were recorded when they attended one of the 22 research centres located across the United Kingdom. Co-morbidity details of these subjects was taken from the self-reported non-cancer illness codes in the UK Biobank (Data-field 20002), which includes self-reported hypertension, myocardial infarction, stroke, diabetes mellitus and arthritis. Arthritis status consists of self-reported rheumatoid arthritis and osteoarthritis.

Statistical analysis

Imputed genotypes of chromosome 16 for all 1474 individuals were used for the association study. Imputation was conducted by UK Biobank using IMPUTE2[54]. As this is a candidate gene association study, we focused on the coding region of *PIEZO1* gene. Variants with imputation quality (Info) score above 0.4 were retained in the analysis. The variants were further filtered based on minor allele frequency (MAF) >5%, missingness >10%, Hardy-Weinberg equilibrium (HWE) $>1x10^{-6}$ and sample missingness >10%. Sixty-five individuals were removed due to missing genotype, leaving 1409 individuals. A total number of 197,730 variants spanning chromosome 16 were eligible for downstream analysis. A further association study was restricted to the self-reported British population from the cohort of 1409 individuals. This derived 1158 individuals who self-reported as British and the filtering parameters were applied as above, resulting in 196,314 variants. Given that the present study is a candidate gene association study, we analysed 334 variants spanning the coding region of *PIEZO1* and the statistical significance was set at $p<0.05$ (0.05/1 gene).

Step 1 of association analysis. The first step was conducted on the association between COVID-19 and *PIEZO1* without adjusting for other underlying disorders. Two general models were performed using logistic regression implemented in PLINK v2.0 (whole genome data analysis toolset) [55]. The first general model was conducted using all the first cohort of individuals tested for COVID-19 (n = 1409), while the second model was limited to those individuals who self-reported as British (n = 1158). The COVID-19 status was used as a categorical variable in the logistic regression analysis. To account for population structure, the principal components were calculated for two cohorts using "--pca approx" implemented in PLINK 2.0 and used as covariates in the logistic regression. The duration of moderate activity was obtained from the UK Biobank (Field-ID 894) and adjusted as a covariate since *PIEZO1* acts as an exercise sensor, mediating optimised redistribution of blood flow to sustain activity[21]. The activity level was determined using a questionnaire on duration of exercise/activity each

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

day. Sex, age, duration of moderate activity and first ten principal components were used as covariates.

Step 2 of association analysis. The step 2 of the analysis focused on British population (*n* = 1158) with adjustment of non-cancer illnesses. The UK Biobank self-reported non-cancer illness codes (Field-ID 20002) were used to define hypertension, myocardial infarction, stroke, diabetes mellitus and arthritis status. Five logistic regression models were used to investigate the association of COVID-19 and *PIEZO1* independently with adjustment of each of the illness. To control for the population structure, first 10 principal components of the 1158 individuals were included as covariates. After logistic regression analysis, ANNOVAR[56] was used to annotate the variants to predict missense variants, frameshift mutations and intronic variants.

Phylogenetic analysis

Protein sequences encoded by the *PIEZO1* gene were obtained from the Ensembl database [57]. Human *PIEZO1* protein (ENSP00000301015) was used as a reference to retrieve orthologous sequences in other species: chimpanzee (ENSPTRP00000014430), rat (ENSRNOP00000069995), mouse (ENSMUSP00000089777), koala (ENSPCIP00000009017), Opossum (ENSMODP00000052005), goldfish (ENSCARP00000072858), guinea pig (ENSCPOP00000030993), Chinese hamster (SCGRP00001016097), sheep (ENSOARP00000014314), Goat (ENSCHIP00000024918), dog (ENSCAFP00000029440), cow (ENSBTAP00000027897), zebrafish (ENSDARP00000101429), tropical clawed frog (ENSXETP00000090443) and electric eel (ENSEEEP00000041564).

The amino acid sequences of these 16 species were aligned using CLUSTAL OMEGA[58]. The resulting alignments were manually curated and trimmed using alignment editor in MEGA 7[59]. Next, the amino acid sequences were subjected to phylogenetic analysis using RAxML[60] based on maximum likelihood algorithm. The phylogenetic tree was generated with 100 rapid bootstrap replicates. The tree was visualised in Fig Tree version 1.4.3[61] and the bootstrap support values labelled on the nodes.

Molecular modelling of PIEZO1

This methodology has been reported previously[20].

DATA AVAILABILITY

The data analysed are available to the registered users through UK Biobank.

CONFLICTS OF INTEREST

The authors declare that they had no conflicts of interest but acknowledge that they were in receipt of research funding as indicated below.

ACKNOWLEDGEMENTS

This research has been conducted using the UK Biobank Resource (Project ID: 42651). The work was supported by a Wellcome Trust Investigator Award to DJB, a BHF Daphne Jackson Fellowship to VD and Leeds Cardiovascular Endowment funds for support of CWC. The study involved use of ARC4, which is part of the High Performance Computing Facility at the University of Leeds. We thank Richard Cubbon for helpful feedback.

REFERENCES

1 Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine* **26**, 450-452, doi:10.1038/s41591-020-0820-9 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2 Bedford, J. *et al.* COVID-19: towards controlling of a pandemic. *Lancet* **395**, 1015-1018, doi:10.1016/S0140-6736(20)30673-5 (2020).

3 WHO. Coronavirus disease (COVID-19) pandemic. *https://www.who.int/emergencies/diseases/novel-coronavirus-2019* (2020).

4 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).

5 Greber, U. F. Virus and Host Mechanics Support Membrane Penetration and Cell Entry. *Journal of Virology* **90**, 3802-3805, doi:10.1128/JVI.02568-15 (2016).

6 de Armas-Rillo, L., Valera, M. S., Marrero-Hernandez, S. & Valenzuela-Fernandez, A. Membrane dynamics associated with viral infection. *Reviews in Medical Virology* **26**, 146-160, doi:10.1002/rmv.1872 (2016).

7 Donate-Macian, P. *et al.* The TRPV4 channel links calcium influx to DDX3X activity and viral infectivity. *Nat Commun* **9**, 2307, doi:10.1038/s41467-018-04776-7 (2018).

8 Hover, S., Foster, B., Barr, J. N. & Mankouri, J. Viral dependence on cellular ion channels - an emerging anti-viral target? *Journal of General Virology* **98**, 345-351, doi:10.1099/jgv.0.000712 (2017).

9 Berridge, M. J., Bootman, M. D. & Roderick, H. L. Calcium signalling: dynamics, homeostasis and remodelling. *Nature Reviews: Molecular Cell Biology* **4**, 517-529, doi:10.1038/nrm1155 (2003).

10 Straus, M. R. *et al.* Ca^{2+} ions promote fusion of Middle East respiratory syndrome coronavirus with host cells and increase infectivity. *Journal of Virology*, doi:10.1128/JVI.00426-20 (2020).

11 Lai, A. L., Millet, J. K., Daniel, S., Freed, J. H. & Whittaker, G. R. The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion Platform that Perturbs Membrane Order in a Calcium-Dependent Manner. *Journal of Molecular Biology* **429**, 3875-3892, doi:10.1016/j.jmb.2017.10.017 (2017).

12 Minakshi, R., Padhan, K., Rehman, S., Hassan, M. I. & Ahmad, F. The SARS Coronavirus 3a protein binds calcium in its cytoplasmic domain. *Virus Research* **191**, 180-183, doi:10.1016/j.virusres.2014.08.001 (2014).

13 Varga, Z. *et al.* Endothelial cell infection and endotheliitis in COVID-19. *Lancet* **395**, 1417-1418, doi:10.1016/S0140-6736(20)30937-5 (2020).

14 Ackermann, M. *et al.* Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *The New England Journal of Medicine*, doi:10.1056/NEJMoa2015432 (2020).

15 Monteil, V. *et al.* Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell* **181**, 905-913 e907, doi:10.1016/j.cell.2020.04.004 (2020).

16 Li, J. *et al.* Piezo1 integration of vascular architecture with physiological force. *Nature* **515**, 279-282, doi:10.1038/nature13701 (2014).

17 Ranade, S. S. *et al.* Piezo1, a mechanically activated ion channel, is required for vascular development in mice. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 10347-10352, doi:10.1073/pnas.1409233111 (2014).

18 Coste, B. *et al.* Piezo1 and Piezo2 are essential components of distinct mechanically activated cation channels. *Science* **330**, 55-60, doi:10.1126/science.1193270 (2010).

19 Guo, Y. R. & MacKinnon, R. Structure-based membrane dome mechanism for Piezo mechanosensitivity. *eLife* **6**, doi:10.7554/eLife.33660 (2017).

20 Chong, J. *et al.* Computational reconstruction of the complete Piezo1 structure reveals a unique footprint and specific lipid interactions. *BioRxiv* **https://doi.org/10.1101/783753** (2019).

21 Rode, B. *et al.* Piezo1 channels sense whole body physical activity to reset cardiovascular homeostasis and enhance performance. *Nat Commun* **8**, 350, doi:10.1038/s41467-017-00429-3 (2017).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

22 Kang, H. *et al.* Piezo1 mediates angiogenesis through activation of MT1-MMP signaling. *American Journal of Physiology: Cell Physiology* **316**, C92-C103, doi:10.1152/ajpcell.00346.2018 (2019).

23 Friedrich, E. E. *et al.* Endothelial cell Piezo1 mediates pressure-induced lung vascular hyperpermeability via disruption of adherens junctions. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 12980-12985, doi:10.1073/pnas.1902165116 (2019).

24 Zhong, M. *et al.* Alveolar Stretch Activation of Endothelial Piezo1 Protects Adherens Junctions and Lung Vascular Barrier. *American Journal of Respiratory Cell and Molecular Biology* **62**, 168-177, doi:10.1165/rcmb.2019-0024OC (2020).

25 Beech, D. J. & Kalli, A. C. Force Sensing by Piezo Channels in Cardiovascular Health and Disease. *Arteriosclerosis, Thrombosis and Vascular Biology* **39**, 2228-2239, doi:10.1161/ATVBAHA.119.313348 (2019).

26 Blythe, N. M. *et al.* Mechanically activated Piezo1 channels of cardiac fibroblasts stimulate p38 mitogen-activated protein kinase activity and interleukin-6 secretion. *Journal of Biological Chemistry* **294**, 17395-17408, doi:10.1074/jbc.RA119.009167 (2019).

27 Cummings, M. J. *et al.* Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. *Lancet*, doi:10.1016/S0140-6736(20)31189-2 (2020).

28 Fotiou, E. *et al.* Novel mutations in PIEZO1 cause an autosomal recessive generalized lymphatic dysplasia with non-immune hydrops fetalis. *Nat Commun* **6**, 8085, doi:10.1038/ncomms9085 (2015).

29 Lukacs, V. *et al.* Impaired PIEZO1 function in patients with a novel autosomal recessive congenital lymphatic dysplasia. *Nat Commun* **6**, 8329, doi:10.1038/ncomms9329 (2015).

30 Fukaya, E. *et al.* Clinical and Genetic Determinants of Varicose Veins. *Circulation* **138**, 2869-2880, doi:10.1161/CIRCULATIONAHA.118.035584 (2018).

31 Zarychanski, R. *et al.* Mutations in the mechanotransduction protein PIEZO1 are associated with hereditary xerocytosis. *Blood* **120**, 1908-1915, doi:10.1182/blood-2012-04-422253 (2012).

32 Ma, S. *et al.* Common PIEZO1 Allele in African Populations Causes RBC Dehydration and Attenuates Plasmodium Infection. *Cell* **173**, 443-455 e412, doi:10.1016/j.cell.2018.02.047 (2018).

33 Nguetse, C. N. *et al.* A common polymorphism in the mechanosensitive ion channel PIEZO1 is associated with protection from severe malaria in humans. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 9074-9081, doi:10.1073/pnas.1919843117 (2020).

34 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).

35 Saotome, K. *et al.* Structure of the mechanically activated ion channel Piezo1. *Nature* **554**, 481-486, doi:10.1038/nature25453 (2018).

36 Niedzwiedz, C. L. *et al.* Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *MedRxiv* **https://doi.org/10.1101/2020.04.22.20075663** (2020).

37 de Lusignan, S. *et al.* Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *The Lancet: Infectious Diseases*, doi:10.1016/S1473-3099(20)30371-6 (2020).

38 UK, O. f. N. S. Coronavirus (COVID-19) related deaths by ethnic group, England and Wales: 2 March 2020 to 10 April 2020. *https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/coronavirusrelateddeathsbyethnicgroupenglandandwales/2march2020to10april2020* (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

39 Gregoret, L. M. & Sauer, R. T. Tolerance of a protein helix to multiple alanine and valine substitutions. *Folding & Design* **3**, 119-126, doi:10.1016/S1359-0278(98)00017-0 (1998).

40 Murthy, S. E., Dubin, A. E. & Patapoutian, A. Piezos thrive under pressure: mechanically activated ion channels in health and disease. *Nature Reviews: Molecular Cell Biology* **18**, 771-783, doi:10.1038/nrm.2017.92 (2017).

41 Evans, E. L. *et al.* RBCs prevent rapid PIEZO1 inactivation and expose slow deactivation as a mechanism of DHS. *Blood*, doi:10.1182/blood.2019004174 (2020).

42 Shi, J. *et al.* Sphingomyelinase disables Piezo1 channel inactivation to enable sustained response to mechanical force. *BioRxiv* **https://doi.org/10.1101/792564** (2019).

43 ClinVar-Miner. *https://clinvarminer.genetics.utah.edu/submissions-by-variant/NM_001142864.4%28PIEZO1%29%3Ac.749T%3EC%20%28p.Val250Ala%29*.

44 Syeda, R. *et al.* Chemical activation of the mechanotransduction channel Piezo1. *eLife* **4**, doi:10.7554/eLife.07369 (2015).

45 Evans, E. L. *et al.* Yoda1 analogue (Dooku1) which antagonizes Yoda1-evoked activation of Piezo1 and aortic relaxation. *British Journal of Pharmacology* **175**, 1744-1759, doi:10.1111/bph.14188 (2018).

46 Ahmed, W. U. R. *et al.* Genome-wide association analysis and replication in 810,625 individuals identifies novel therapeutic targets for varicose veins. *BioRxiv* **https://doi.org/10.1101/2020.05.14.095653** (2020).

47 Millet, J. K. & Whittaker, G. R. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology* **517**, 3-8, doi:10.1016/j.virol.2017.12.015 (2018).

48 Nieto-Torres, J. L. *et al.* Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology* **485**, 330-339, doi:10.1016/j.virol.2015.08.010 (2015).

49 Yue, Y. *et al.* SARS-Coronavirus Open Reading Frame-3a drives multimodal necrotic cell death. *Cell Death & Disease* **9**, 904, doi:10.1038/s41419-018-0917-y (2018).

50 Dalmasso, C. *et al.* Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study. *PloS one* **3**, e3907, doi:10.1371/journal.pone.0003907 (2008).

51 Solis, A. G. *et al.* Mechanosensation of cyclical force by PIEZO1 is essential for innate immunity. *Nature* **573**, 69-74, doi:10.1038/s41586-019-1485-8 (2019).

52 Sargin, G., Yavasoglu, S. I. & Yavasoglu, I. Is Coronavirus Disease 2019 (COVID-19) seen less in countries more exposed to Malaria? *Medical Hypotheses* **140**, 109756, doi:10.1016/j.mehy.2020.109756 (2020).

53 Napoli, P. E. & Nioi, M. Global Spread of Coronavirus Disease 2019 and Malaria: An Epidemiological Paradox in the Early Stage of A Pandemic. *Journal of Clinical Medicine* **9**, doi:10.3390/jcm9041138 (2020).

54 Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457-470, doi:10.1534/g3.111.001198 (2011).

55 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

56 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164, doi:10.1093/nar/gkq603 (2010).

57 Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688, doi:10.1093/nar/gkz966 (2020).

58 Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145, doi:10.1002/pro.3290 (2018).

59 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.01.20119651; this version posted June 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

60 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

61 Rambaut, A. FigTree, a graphical viewer of phylogenetic trees. (2007).