Article

EFCAB4B (CRACR2A/Rab46) Genetic Variants Associated with COVID-19 Fatality

Dapeng Wang [1,2], Sabina D. Wiktor [3], Chew W. Cheng [3], Katie J. Simmons [3], Ashley Money [3], Lucia Pedicini [3], Asya Carlton [4], Alexander L. Breeze [4] and Lynn McKeown [3,*]

[1] LeedsOmics, University of Leeds, Leeds LS2 9JT, UK; dapeng.wang@imperial.ac.uk
[2] National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK
[3] Leeds Institute of Cardiovascular and Metabolic Medicine, Faculty of Medicine and Health, University of Leeds, Leeds LS2 9JT, UK; c.w.cheng@leeds.ac.uk (C.W.C.); k.j.simmons@leeds.ac.uk (K.J.S.)
[4] Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK
[*] Correspondence: l.mckeown@leeds.ac.uk

Abstract: The coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has resulted in more than 692 million cases worldwide and nearly 7 million deaths (August 2023). Severe COVID-19 is characterised in part by vascular thrombosis and a cytokine storm due to increased plasma concentrations of pro-thrombotic proteins such as von Willebrand factor and cytokines secreted from endothelial and T-cells. *EFCAB4B* is a gene that encodes for two proteins (CRACR2A and Rab46) that play important roles in endothelial and T-cell secretion. In this study, using patient data recorded in the UK Biobank, we demonstrate the importance of variants in the *EFCAB4B* genetic sequence with COVID-19 fatality. Using logistic regression analysis, we determined that three single-nucleotide polymorphisms (SNPs) in the gene cause missense variations in CRACR2A and Rab46, which are associated with COVID-19 fatality (rs9788233: $p = 0.004$, odds ratio = 1.511; rs17836273: $p = 0.012$, odds ratio = 1.433; rs36030417: $p = 0.013$, odds ratio = 1.393). All three SNPs cause changes in amino acid residues that are highly conserved across species, indicating their importance in protein structure and function. Two SNPs, rs17836273 (A98T) and rs36030417 (H212Q), cause amino acid substitutions in important functional domains: the EF-hand and coiled-coil domain, respectively. Molecular modelling shows minimal impact by the substitution of threonine at position 98 on the structure of the EF-hand. Since Rab46 is a GTPase that regulates both endothelial cell secretion and T-cell signalling, these missense variants may play a role in the molecular mechanisms underlying the thrombotic and inflammatory characteristics observed in patients with severe COVID-19 outcomes.

Keywords: COVID-19; fatality; CRACR2A; Rab46; *EFCAB4B*; SNPs

1. Introduction

Coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) triggered a global pandemic in 2020 [1]. Although COVID-19 affects the respiratory system, cardiovascular complications, including abnormal clotting, thrombosis and microvascular injury [2,3], are also part of the disease spectrum and associated with increased morbidity and mortality. Several studies have suggested that SARS-CoV-2 can directly infect the endothelial cells (ECs) that line the blood vessels: indeed, endothelial dysfunction is found in severe cases of COVID-19 [4–6]. Activation of ECs triggers exocytosis of pro-thrombotic mediators including von Willebrand factor (vWF), angiopoietin-2 and P-selectin, all of which are elevated in the serum of severely affected patients and are significantly associated with in-hospital fatality [6–11]. Hospitalised patients also present with hyper-inflammation, having excessive levels of cytokines (cytokine storm) and other inflammatory markers that are characteristic of increased EC

activity and lymphocyte activity [12]. These data suggest that the inappropriate degranulation of both ECs and lymphocytes contributes to the pro-thrombotic, pro-inflammatory and excessive cytokine environment observed in severe COVID-19 cases. We suggest that genetic differences in a common molecular mechanism underlying degranulation could confer susceptibility to both the vascular and immune abnormalities induced by COVID-19 with the subsequent increased risk of death.

In ECs, pro-thrombotic and pro-inflammatory proteins are stored in specialised vesicles called Weibel–Palade bodies (WPBs) [13]. WPBs are endothelial cell-specific organelles that, upon the release of contents, control important vascular events such as haemostasis, inflammation, angiogenesis and vascular tone. During vascular injury, WPBs undergo rapid degranulation and eject cargo into the vessel lumen including the following: vWF filaments to initiate platelet plug formation [14]; P-selectin to attract leukocytes [15] and angiopoietin-2 to induce EC migration [16]. Since the release of WPB cargo into the vessel lumen is highly regulated, there must be molecular changes that underlie the inappropriately raised plasma levels observed in COVID-19 patients. Excessive cargo release may be evoked by the direct viral activation of ECs or an increase in cytokines acting on ECs; however, the thrombo-inflammatory aberrations are not apparent in all infected patients (even those with high viral load). This indicates that there is either a pre-existing deterioration of the endothelium (co-morbidity), or patients may have an increased susceptibility to severe COVID-19 due to their genetic makeup, particularly in the genes that encode for proteins that play a role in regulating the release of WPBs from ECs. We study the role of a novel Rab GTPase (Rab46: CRACR2A-L; Ensembl CRACR2A-203; CRACR2A-a [17]) in EC degranulation. We have previously reported that Rab46 is necessary for anchoring a subpopulation of WPBs to the microtubule-organising centre in response to histamine signalling [18]. This stimuli-coupled trafficking inhibits the release of WPB cargo that is superfluous to histamine function and thus prevents the full thrombotic response evoked by vessel injury. Evidence of Rab46 function beyond EC biology is suggested by a function for Rab46 in immune cells such as β- and T-lymphocytes and mast cells [19]. T-cell receptor stimulation evokes Rab46-dependent vesicle translocation to the immunological synapse and the activation of the vav1/JNK signalling pathways necessary for secretion [19]. Indeed, we have reported that a patient with biallelic Rab46/CRACR2A mutations has a compromised immune system due to reduced T-cell signalling with subsequent defective cytokine production [20]. Since Rab46 function may regulate a secretory path common to various cell types involved in COVID-19 severity, it is important to determine any associations between variants in the gene that encodes for Rab46 with COVID-19-related susceptibility and fatality.

Rab46 (732 amino acids) is one of two functional isoforms from a possible six transcripts of the gene *EFCAB4B* and the only isoform expressed in ECs. CRACR2A (CRACR2A-S; Ensembl CRACR2A-201; CRACR2A-c) is a shorter (395 amino acids), non-Rab isoform expressed in T-cells and regulates store-operated calcium entry [21] and T-cell signalling [22] (Figure 1). The alignment of the amino acid sequences of Rab46 and CRACR2A reveals that the *N*-terminal components are identical, but Rab46 contains a distinct long C-terminus containing an extra Rab GTPase domain.
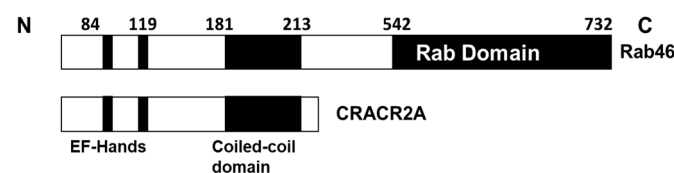

Figure 1. Schematics of the functional domains in the *EFCAB4B* isoforms Rab46 and CRACR2A.

Recently, a longitudinal analysis of patients who recovered from mild COVID-19 reported decreased CRACR2A/Rab46 levels in the switch from an inflammatory to a wound healing response for the recovery phase of the infection [23]. Moreover, a study identified

CRACR2A/Rab46 proteins as being significantly raised in the blood of patients with long COVID compared to those who had acute infections [24]. Therefore, we hypothesise that genetic variants in *EFCAB4B* may be instrumental in a subset of COVID-19 fatalities. Here, we used a candidate gene association study to determine the links between COVID-19 fatality and variants in the *EFCAB4B* gene using COVID-19 data recorded in the UK Biobank prior to the vaccination programme.

2. Materials and Methods

2.1. Demographic Data

The UK Biobank (application ID: 42651), from which we obtained data for use in this study, collected data on around 500,000 individuals in the UK between 2006 and 2010 [25]. These individuals consented to the UK Biobank obtaining their genomic data, health records and information regarding demographics and lifestyle. The UK Biobank released the data on 10,118 individuals prior to the vaccination program, who self-identified as British and had been tested for COVID-19, with their associated outcomes also recorded (Table 1). Co-morbidity data were also available for those individuals, including hypertension, diabetes, asthma, arthritis, stroke and myocardial infarction. The chi-squared test was used for categorical variables, and the Kruskal–Wallis rank sum test was used for continuous variables in the statistical analysis for the comparison among the three groups shown in Table 1.

Table 1. Demographics of UK Biobank patients used in this study. Statistical tests were performed for comparing the three groups: COVID-19 negative, Fatal and Non-fatal. Chi-squared test was used for categorical variables and Kruskal-Wallis rank sum test was used for continuous variables.

Characteristic	Total N	COVID-19 Negative	COVID-19 Positive			*p*-Value
			Fatal and Non-Fatal	Fatal	Non-Fatal	
Number of Individuals	10,118	8853	1265	194	1071	
Age	70 ± 8.3	70.1 ± 8.2	68.9 ± 9.2	75.1 ± 5.7	67.7 ± 9.2	1.38×10^{-30}
Sex						
Male	4972	4309	663	129	534	5.09×10^{-6}
Female	5146	4544	602	65	537	
Co-Morbidity						
Myocardial Infarction	429	375	54	13	41	0.19
Stroke	256	221	35	6	29	0.81
Arthritis	1477	1325	152	37	115	2.16×10^{-4}
Diabetes Mellitus	731	627	104	27	77	1.34×10^{-3}
Hypertension	3436	3010	426	86	340	2.95×10^{-3}
Asthma	1380	1222	158	17	141	0.12

2.2. Analysis of EFCAB4B Variants Associated with COVID-19 Outcomes

Associations between *EFCAB4B* variants and COVID-19 fatality were determined by logistic regression analysis, using an additive method. The genotype data were mainly handled with the PLINK 2.0 analytical framework [26]. The initial screening process generated the samples from those with a British ethnic background (coding: 1001) that have COVID-19 test data and the variants with imputation info score > 0.4 (Table S1). Further filtering was carried out using —geno 0.1, —maf 0.01 and —hwe 0.000001 for variants and using —mind 0.1 for samples. In addition, only the variants residing on the *EFCAB4B* gene were retained according to the start coordinate (3714799) and end coordinate (3874985) on human chromosome 12 based on the genome assembly version GRCh37.

The three groups were defined as follows. The fatal group comprises individuals who tested positive for COVID-19 and who were also labelled with COVID-19 deaths (coding: U071). The non-fatal group consists of individuals who tested positive for COVID-

19 with no records in the cause of death data. The negative group is made up of those who tested negative for COVID-19 that are not identified in the cause of death data. To test the association of genotypes with phenotypes, the logistic regression was performed for case/control phenotypes accounting for covariates including sex, age and the first ten principal components. This enables the identification of whether the presence of variants is a risk factor for COVID-19 fatality by comparing the three groups. We created three comparative cohorts: (1) Non-fatal vs. negative: Patients who tested positive for COVID-19 and survived (case) as compared to those who had not been diagnosed with COVID-19 (control); (2) Fatal vs. negative: Patients who tested positive for COVID-19 and died (case) as compared to those who had not been diagnosed with COVID-19 (control); (3) Fatal vs. non-fatal: Patients who had tested positive for COVID-19 and died (case) as compared to those who survived (control). The first two comparative cohorts were used to determine whether variants are significantly associated with disease susceptibility (cohort 1) or fatality (cohort 2) as compared to a negative population. To determine which variants are exclusively associated with fatality, we compared fatal vs. non-fatal (cohort 3).

For the adjustment of illnesses, an additional covariate was added respectively for the six co-morbidities such as hypertension, myocardial infarction, stroke, diabetes mellitus, arthritis and asthma obtained from non-cancer illness code. Ensembl Variant Effect Predictor (VEP) was used to predict and annotate the effect and impact of the variant on the gene functions from the human genome (GRCh37). The SNPs identified in Table 2 were those that caused missense variations and were nominally statistically significant ($p < 0.05$). Corrections for multiple testing (False Discover Rate: FDR) are included in the Supplementary File (Tables S3–S5).

Table 2. Characterisation of three missense Rab46/CRACR2A SNPs.

SNP ID	MAF (ALFA)	REF	ALT	Type	Amino Acid Change
rs9788233	0.119	T	C	substitution	R7G
rs17836273	0.124	C	T	substitution	A98T
rs36030417	0.145	A	T	substitution	H212Q

MAF = minor allele frequency. REF = reference allele. ALT = alternate allele.

2.3. Conservation of Rab46 Protein Sequence across Different Species

The Rab46 sequence conservation diagram was generated using Basic Local Alignment Search Tool (BLAST [27]) and WebLogo 3 [28]. The amino acid sequence of Rab46 (NP_001138430.1) was searched against the protein sequence database in BLAST (BLASTP) and aligned with the identified 100 similar protein sequences (percent sequence identity > 80%) from 66 different species using the COBALT multiple alignment tool. The generated sequence alignment enabled the calculation of residue conservation scores. WebLogo 3 [28,29] was used to visualise the level of amino acid sequence conservation around residues for which missense variations were detected.

2.4. Rab46 Structural Analysis

The predicted AlphaFold [30,31] structure of full-length human Rab46 and the crystal structure of the EF-hand domain (PDB: 6PSD) were pre-processed and energy-minimised using the Protein Preparation Wizard (Maestro software, Release 2020-1, Glide, Schrödinger, LLC, New York, NY, USA, 2020).

The effect of the A98T variation was modelled using the available crystal structure of the EF-hand domain of Rab46. Alanine residue at position 98 was mutated to threonine using the 'Mutate Residue' tool in the Maestro graphical user interface (Maestro software, Release 2020-1, Glide, Schrödinger, LLC, New York, NY, USA, 2020). Side chains of amino acids within 5 Å from the introduced variation were minimised using the OPLS3 forcefield [32]. The alignment of the wild-type and mutant EF-hand domain structures was visualised using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger,

LLC) and changes to the structure visualised using Pymol and the Maestro graphical user interface.

3. Results

3.1. Demographics

Here, we analysed genomic data held in the UK Biobank from 10,118 subjects, who had self-identified as British. Of these, 50.86% were female and 49.14% males, with a mean age of 70 ± 8.3 years (means ± SD). In total, 1265 patients were diagnosed COVID-19 positive (Table 1). The positive cases were further categorised into fatal and non-fatal, with fatal indicating the individuals who died due to COVID-19 based on data obtained from the National Death Registries Database. Among those who tested positive, 194 (15.34%) were reported to have died due to COVID-19 (fatal), while 1071 (84.66%) were non-fatal (Table 1). In addition, because any determined associations between *EFCAB4B* variants and COVID-19 fatality could be due to the variant predisposing an individual to a COVID-19-susceptible health condition, we evaluated the co-morbidities of these individuals. Of the 1265 COVID-19-positive individuals, 426 (33.68%) self-reported hypertension, 54 (4.27%) myocardial infarction, 35 (2.77%) stroke, 104 (8.22%) diabetes mellitus, 152 (12.02%) arthritis and 158 (12.49%) asthma.

3.2. Variants in the EFCAB4B Gene Are Associated with Susceptibility to COVID-19 Infection and COVID-19 Fatality

The Venn diagram shown in Figure 2 depicts the overlap of the variants significantly ($p < 0.05$) associated with each comparative cohort (Tables S3–S5). A total of 128 variants are associated with COVID-19 infection (both fatal and non-fatal) compared with the non-infected population, and 106 variants are associated with COVID-19 fatality when compared to non-infected populations (cohort 2). Among them, the overlap between cohort 2 and cohort 3 indicates that 54 variants are distinctly associated with fatality and not due to disease susceptibility. Herein, we specifically explored the variants associated with fatality or susceptibility that affect the translation of the *EFCAB4B* gene into its relevant proteins by inducing missense variations; however, the identified intronic variants merit further analysis and could play an important regulatory role in gene expression or isoform levels.
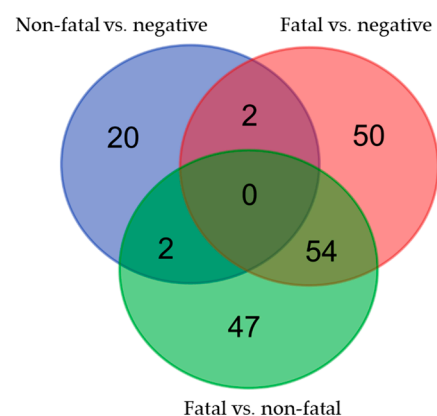

Figure 2. A Venn Diagram of Variants Associated with the Test Cases in Each of the Three Cohorts: (1) non-fatal vs. negative (blue); (2) fatal vs. negative (red) and (3) fatal vs. non-fatal (green).

3.3. EFCAB4B Missense Variants Associated with COVID-19

Three SNPs with a minor allele frequency (MAF) > 0.05 were identified as causing missense variations in both CRACR2A and Rab46 (Table 2, Figure 3). We determined these SNPs as being significantly associated with COVID-19 fatality in cohort 2 (Table 3). The results showed the following: (1) the C allele of rs9788233 is significantly associated with fatality with an OR of 1.511 ($p = 0.004$); (2) the T allele of rs17836273 has an OR of 1.433 ($p = 0.012$) and (3) the T allele of rs36030417 has an OR of 1.393 ($p = 0.013$) (Table 3).

There variants are nominally significant ($p < 0.05$) but not significant after correcting for multiple testing (Tables S3–S5). As a further measure of association with COVID-19 fatality, we also analysed associations between fatal versus non-fatal (Table 3). In these analyses, rs17836273 and rs36030417 displayed $p = 0.099$ and $p = 0.054$, respectively; however, due to the significant associations between these SNPs and COVID-19 susceptibility being reported in the REGENERON study (meta-analyses of data collected in UK Biobank, ANCESTRYDNA and TopMed: [33,34]), and the location of these variants in important domains that regulate protein function, we included them in further analyses. Firstly, since COVID-19 susceptibility and fatality may be due to co-morbidities, we adjusted the data to account for six underlying disorders known as risk factors for COVID-19: hypertension, myocardial infarction, stroke, diabetes mellitus, arthritis and asthma (Table 4). Consistently across six models of logistic regression, the results supported the relevance of the same three missense variants to COVID-19 susceptibility and fatality.
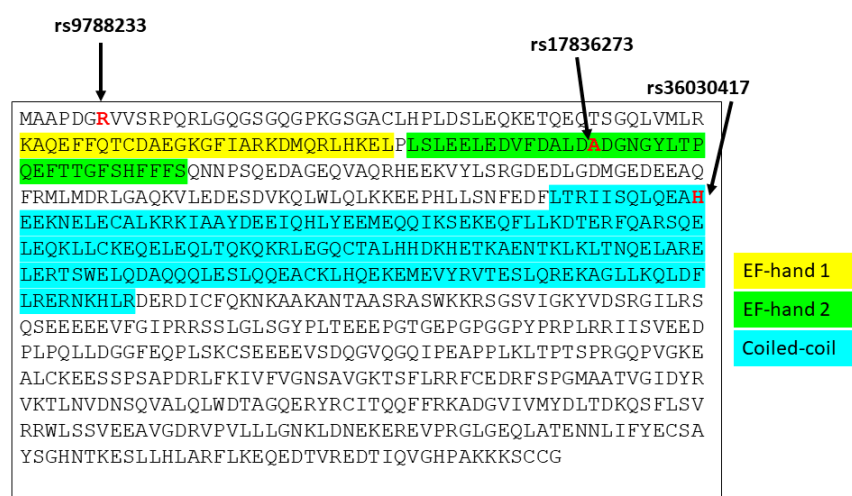

Figure 3. The human amino acid sequence of Rab46 indicating the position of the SNPs (highlighted in red) significantly associated with COVID-19 fatality within the highlighted functional domains.

Table 3. Association analysis of three missense Rab46/CRACR2A SNPs with COVID-19 outcomes.

SNP ID	Effect Allele	Non-Fatal vs. Negative *p*-Value/OR	Fatal vs. Negative *p*-Value/OR	Fatal vs. Non-Fatal *p*-Value/OR
rs9788233	C	-	0.004/1.511	0.021/1.487
rs17836273	T	-	0.012/1.433	0.099/1.314
rs36030417	T	-	0.013/1.393	0.054/1.351

OR = odds ratio. - indicates no significant association found.

Table 4. COVID-19 outcome and Rab46 missense SNPs association analysis after adjustment for co-morbidities.

Patient Cohort	SNP ID	Effect Allele	Hypertension *p*-Value/OR	Diabetes *p*-Value/OR	Asthma *p*-Value/OR	Arthritis *p*-Value/OR	Stroke *p*-Value/OR	Myocardial Infarction *p*-Value/OR
Fatal vs. negative	rs9788233	C	0.004/1.512	0.004/1.516	0.005/1.505	0.004/1.51	0.004/1.511	0.004/1.511
	rs17836273	T	0.012/1.435	0.011/1.44	0.014/1.426	0.013/1.433	0.012/1.433	0.012/1.433
	rs36030417	T	0.013/1.395	0.012/1.398	0.015/1.386	0.013/1.394	0.013/1.393	0.013/1.394
Fatal vs. non-fatal	rs9788233	C	0.022/1.481	0.018/1.498	0.022/1.482	0.024/1.474	0.022/1.483	0.021/1.487
	rs17836273	T	0.105/1.309	0.089/1.325	0.102/1.311	0.117/1.297	0.104/1.309	0.1/1.313
	rs36030417	T	0.054/1.352	0.05/1.359	0.056/1.35	0.062/1.34	0.053/1.353	0.055/1.351

OR = odds ratio.

3.4. EFCAB4B Missense Variants Affect Highly Conserved Residues

These three missense SNPs induce amino acid substitutions in both the CRACR2A and Rab46 protein when translated from the *EFCAB4B* gene (Table 2, Figure 3). The arginine to glycine (R7G) change induced by rs9788233 occurs at the N-terminus in a structurally undefined region. The substitutions incurred by both rs17836273 and rs36030417 (alanine to threonine A98T and histidine to glutamine H212Q, respectively) occur in functionally important domains found in both CRACR2A and Rab46 (the canonical motif of the EF-hand domain and the coiled-coil domain). To determine the importance of these amino acid residues, we explored their evolutionary conservation (Figure 4). The conservation of Rab46 across 100 different sequences and 66 species was obtained through conducting BLAST sequence alignments of the Rab46 protein sequence. From the multiple alignments, it was found that arginine (rs9788233) in a domain of unknown function was conserved across 80% of species. Other common amino acids found in the 20% of non-conserved species were lysine and asparagine. Alanine (rs17836273) at position 98 in the second EF-hand motif was highly conserved (98%), and histidine (rs36030417) located in the coiled-coil domain was the most conserved (99%). The high level of conservation of these residues suggests that these amino acids are critical to the function of Rab46 and CRACR2A.
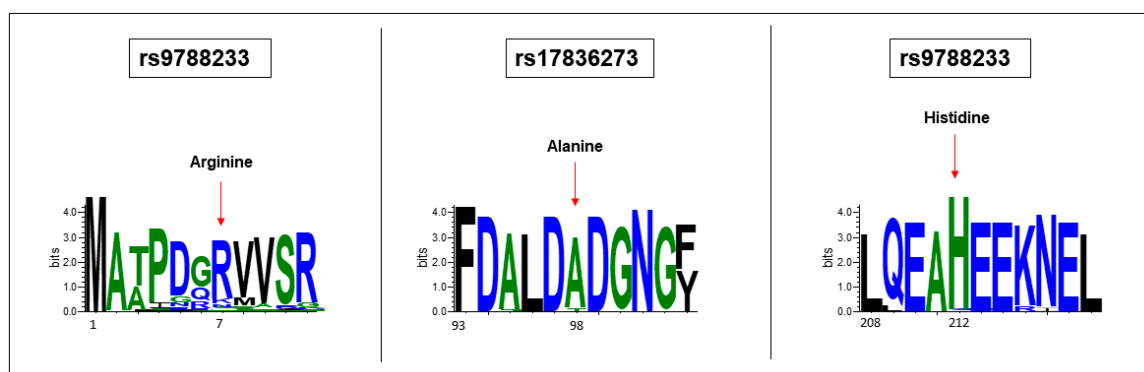

Figure 4. Sequence logo plots depicting the conservation of the specified amino acid residues compared across 100 sequences and 66 species demonstrating the high conservation of the three SNPs.

3.5. Prediction of Missense Variants on Rab46 Structure

The structure of Rab46 is unknown, but we sought further insight into the role of the missense variants in protein function by the analysis of an AlphaFold prediction of the structure of full-length Rab46 and the available crystal structure of the EF-hand domain (Figure 5). The AlphaFold prediction, whilst speculative, suggests that the EF-hand and the coiled-coil domains are of high confidence, and here, we suggest that these residues likely have solvent-exposed side chains (highlighted in yellow).

We further explored the A98T variation through the availability of the crystal structure for the EF-hand domain (PDB: 6PSD). To understand the possible impact of the threonine residue that is substituted for alanine (A98T) in rs17836273, we used molecular modelling to overlay the crystal structure of the wild-type (WT) alanine-containing variant (blue) with the threonine-substituted variant (cream: Figure 6). Rab46 contains two EF-hand motifs; however, the crystal structure and Ca^{2+} binding data predict that only the second EF-hand coordinates Ca^{2+} (Figure 6). The aspartate/alanine/aspartate (DAD: the WT structure) motif in this second EF-hand is necessary for coordinating calcium [35]. Overlaying the WT crystal structure with the modelled structure of the A98T variant (shown in cream) reveals only minor structural changes within the binding domain, entailing a slight (<1 Å) shift in the D97-G100 loop.
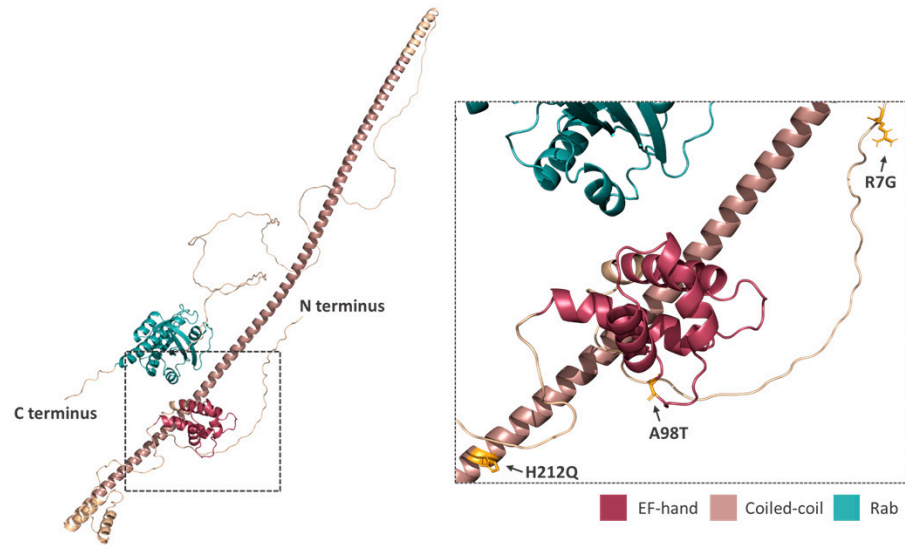
Figure 5. The prediction of the position and orientation of the three missense SNPs in an AlphaFold model of a full-length Rab46 structure. The box indicated by the dotted line in the full-length structure is enlarged in order to demonstrate the position and orientation of the side chains of the three missense SNPs: R7G; A98T and H212Q.
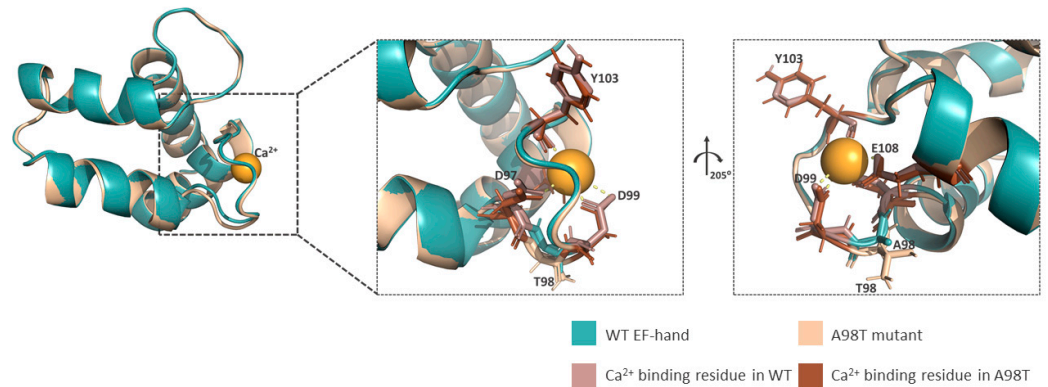

Figure 6. The modelling of the effect of A98T substitution on the EF-hand domain of Rab46. Right: the crystal structure of the wild-type (WT) and a modelled structure of the mutant (A98T) EF-hand domain aligned using PyMOL. Left: expanded boxes demonstrating subtle changes within the calcium binding loop.

In summary, we have described three missense variants in the *EFCAB4B* gene that are associated with fatality in patients with COVID-19, of which two could cause changes to protein structure and function.

4. Discussion

Here, we performed logistic regression analysis using recorded COVID-19 outcomes from patient data deposited in the UK Biobank, prior to the vaccination programme, and links between known *EFCAB4B* variants. Three missense variants (rs9788233, rs17836273, rs36030417) that caused substitutions in the translated protein were significantly associated with COVID-19 fatality when compared to a negative population. This suggests that these SNPs induce genetic variations in the *EFCAB4B* gene that are associated with COVID-19 fatality. As a more predictive measure of fatality, we also analysed associations between fatal versus non-fatal (cohort 3). rs17836273 and rs36030417 did not show significance at $p < 0.05$ in this comparison, therefore suggesting these SNPs are associated with susceptibility to COVID-19 infection. We recognise that a potential limitation of our study is the relatively small size of the COVID-19 data set, especially as the higher p values observed here could

be due to the small sampling size in cohort 3. An analysis of additional data as they become available will be important for a deeper understanding and exploration of *EFCAB4B*, especially in different populations and ethnicities. However, the association of these variants with COVID-19 susceptibility still validated them to be included in this analysis. A synonymous variant, rs17697920, was also found to be significantly associated with COVID-19 fatality but not included in our analysis as CTG-CTA both encode for leucine. In addition, intronic variants were observed in all comparative groups. Whilst further analysis has focused on the missense variants, many of the intronic variants identified as being significantly associated with both COVID-19 susceptibility and fatality could be of interest, especially if they cause changes in expression levels or isoform splicing. We will specifically investigate these variants in our future studies when we explore the regulation of *EFCAB4B* splicing and determine how the expression levels of each isoform (CRACR2A/Rab46) contribute to health and disease.

Two missense variants, rs17836273 and rs36030417, are highly conserved (98% and 99%, respectively) and are located in the N-terminus of both the CRACR2A and Rab46 protein in domains that have roles in protein function. Variant rs36030417 causes a histidine to glutamine switch in a region that is predicted to be a coiled-coil domain. Coiled-coils have several important roles in proteins, but they are mostly important for facilitating protein–protein interactions. Histidine and glutamine are fairly conserved in size, which is important for residues that have to be packed inside an alpha helix; however, the imidazole ring on histidine could be important for metal coordination which is often used in coiled-coils for the reinforcement of an oligomerised protein complex [36]. Particularly, it is suggested by the predicted AlphaFold structure that the amino acid at position 212 orientates outwardly and therefore would be more important in molecular interactions than helix stability. Both Rab46 and CRACR2A are suggested to act as multimers, and therefore, the biochemical analysis of this variant is necessary to determine the impact it has on this function.

Variant rs17836273 is located in the EF-hand domain of both Rab46 and CRACR2A. EF-hands consist of a helix–loop–helix motif, and the binding of calcium to the loop region causes a change in the EF-hand conformation, which is propagated through a protein to enable function. The second EF-hand of Rab46 consists of a canonical calcium binding motif of 12 amino acids (DADGNGYLTPQE). Amino acids in positions 1, 3, 5, 7, 9 and 12 in this motif are important in the coordination of the calcium ion [37]. The A98T substitution conferred by this SNP is a fairly conserved alanine to threonine variation, and it is located in position 2 of the motif, which is thought to be a changeable site. Overlaying the crystal structure of the EF-hand domain and an in silico model of the A98T form, we observe that the shape of the calcium binding loop is minimally altered by the threonine substitution, and Ca^{2+} binding would appear unaffected. On the available structural and modelling evidence at the domain level, the effect of the variation is thus currently unexplained. The effect of A98T will need further biochemical and biophysical (ITC, NMR) analysis to determine the impact on both calcium binding and cell physiology; it is possible that its effect is manifested at the level of domain interactions rather than individual domain function.

Similar coiled-coil and EF-hand-containing proteins (e.g., STIM1) have variation in these domains that cause disease; for example, an activating variation in the STIM1 coiled-coil region is associated with Stormoken syndrome [38], whilst an EF-hand variation causes tubular aggregate myopathy [39]. We do not know if these variants in CRACR2A/Rab46 evoke activation or inactivation. We know that in endothelial cells, Rab46 acts as a brake in response to certain stimuli to prevent the total release of WPB contents [18]. Therefore, we could predict that the inhibition of this function would lead to an increase in the release of pro-thrombotic and pro-inflammatory cargo thus leading to some of the clinical features observed in severe cases of COVID-19. In T-cells, Rab46 is important for T-cell signalling [22]; moreover, we have shown that a patient who has biallelic mutations in the gene *EFCAB4B* (double mutation in allele1: R144G and E300*; allele 2 E278D), so that they no longer express full-length Rab46 or CRACR2A, exhibits reduced cytokine

expression due to decreased calcium influx and JNK signalling [20]. This suggests that the overexpression or over-activation of Rab46 in T-cells could lead to the increase in cytokine release as observed in patients with cytokine storm reactions to COVID-19.

A role for Rab46/CRACR2A in the immune response to COVID-19 infection is indicated in a study by Furuyama et al. [40]. The intramuscular injection of a SARS CoV-2 spike-containing vaccine in Rhesus macaques resulted in a cellular and humoral immune response and provided protection from COVID-19-driven pneumonia compared to intranasal administration. The transcriptional analysis of differentially expressed genes from the lung tissues of infected animals demonstrated the upregulation of Rab46/CRACR2A in the protected monkeys. Moreover, in a study by Talla et al., the characterisation of signals associated with recovery in individuals with mild COVID-19 demonstrated a reduced expression of inflammatory proteins, including Rab46/CRACR2A, after the early immune response [23]. The indication that CRACR2A impacts disease severity due to its role in the immune response is also supported by the function of CRACR2A in T-cells, where CRACR2A is necessary for store-operated calcium entry (SOCE) [21]. This is of particular importance as SOCE impacts COVID-19 susceptibility [41] through the regulation of interferon-dependent immunity [42], a system where the loss of function and rare variants are highly associated (OR > 27) with life-threatening COVID-19 infection [43], particularly in younger patients [44]. More recently, Patel et al. (24) identified CRACR2A/Rab46 as a protein found in patients with long COVID. These studies indicate that regulating the expression of Rab46 and/or CRACR2A is important for both the initial response and recovery phases of COVID-19 infection.

To understand the importance of the identified *EFCAB4B* missense variants being associated with COVID-19 fatality requires further structural analysis and longer-term studies using appropriate animal models. Particularly, the generation of CRISPR-Cas9 endothelial/immune cells and mouse models expressing either the missense or some of the intronic variants will provide a valuable understanding of the mechanisms underlying CRACR2A/Rab46 function and how these proteins contribute to COVID-19 infectivity or fatality. Exploring how genetic variants influence disease also informs us of the physiological pathways that play roles in disease susceptibility or severity, and this is of particular importance when we consider that age and sex are risk factors for COVID-19. Whilst *EFCAB4B* is not sex-linked, an understanding of how the immune system and endothelium contribute to disease will provide novel targets for therapeutic treatments. We also need to be aware of the limitations of this study because the reported data are from a relatively small sample size in an aged population prior to the rollout of a vaccination programme in the UK. Moreover, there may be population differences in the allele frequencies, and due to the number of samples, we had to restrict our analysis to those who self-identified as British. The obtained results would require further validation in larger cohorts with different ethnicities and geographical regions.

5. Conclusions

COVID-19 has been generally accepted as a multigenic and multifactorial disease with different determinants. Although the understanding of the pathophysiology of COVID-19 is improving, the prognosis is poor in many patients with severe disease. The identification of the genetic variants associated with susceptibility to COVID-19 infection or fatality is the key to better understanding the pathological mechanisms underlying an individual's increased risk for COVID-19 infection or severity. The results of the present study, despite their intrinsic limitations, showed that SNPs in *EFCAB4B* could be associated with a higher risk for COVID-19 infection or fatality. From a clinical perspective, further studies of this gene with regards to COVID-19 infection, or indeed other viral pathologies, may present *EFCAB4B* as a potential biomarker and thereby inform of clinical risk management and, eventually, personalised treatment plans.

References

1. Atzrodt, C.L.; Maknojia, I.; McCarthy, R.D.P.; Oldfield, T.M.; Po, J.; Ta, K.T.L.; Stepp, H.E.; Clements, T.P. A Guide to COVID-19: A global pandemic caused by the novel coronavirus SARS-CoV-2. *FEBS J.* **2020**, *287*, 3633–3650. [CrossRef]
2. McFadyen, J.D.; Stevens, H.; Peter, K. The Emerging Threat of (Micro)Thrombosis in COVID-19 and Its Therapeutic Implications. *Circ. Res.* **2020**, *127*, 571–587. [CrossRef]
3. Wazny, V.; Siau, A.; Wu, K.X.; Cheung, C. Vascular underpinning of COVID-19. *Open Biol.* **2020**, *10*, 200208. [CrossRef]
4. Nägele, M.P.; Haubner, B.; Tanner, F.C.; Ruschitzka, F.; Flammer, A.J. Endothelial dysfunction in COVID-19: Current findings and therapeutic implications. *Atherosclerosis* **2020**, *314*, 58–62. [CrossRef]
5. Zhang, J.; Tecson, K.M.; McCullough, P.A. Endothelial dysfunction contributes to COVID-19-associated vascular inflammation and coagulopathy. *Rev. Cardiovasc. Med.* **2020**, *21*, 315–319. [CrossRef]
6. Goshua, G.; Pine, A.B.; Meizlish, M.L.; Chang, C.H.; Zhang, H.; Bahel, P.; Baluha, A.; Bar, N.; Bona, R.D.; Burns, A.J.; et al. Endotheliopathy in COVID-19-associated coagulopathy: Evidence from a single-centre, cross-sectional study. *Lancet Haematol.* **2020**, *7*, e575–e582. [CrossRef]
7. Villa, E.; Critelli, R.; Lasagni, S.; Melegari, A.; Curatolo, A.; Celsa, C.; Romagnoli, D.; Melegari, G.; Pivetti, A.; Di Marco, L.; et al. Dynamic angiopoietin-2 assessment predicts survival and chronic course in hospitalized patients with COVID-19. *Blood Adv.* **2021**, *5*, 662–673. [CrossRef]
8. Fenyves, B.G.; Mehta, A.; Kays, K.R.; Beakes, C.; Margolin, J.; Goldberg, M.B.; Hacohen, N.; Filbin, M.R. Plasma P-selectin is an early marker of thromboembolism in COVID-19. *Am. J. Hematol.* **2021**, *96*, E468–E471. [CrossRef]
9. Yatim, N.; Boussier, J.; Chocron, R.; Hadjadj, J.; Philippe, A.; Gendron, N.; Barnabei, L.; Charbit, B.; Szwebel, T.A.; Carlier, N.; et al. Platelet activation in critically ill COVID-19 patients. *Ann. Intensive Care* **2021**, *11*, 113. [CrossRef]
10. Cotter, A.H.; Yang, S.T.; Shafi, H.; Cotter, T.M.; Palmer-Toy, D.E. Elevated von Willebrand Factor Antigen Is an Early Predictor of Mortality and Prolonged Length of Stay for Coronavirus Disease 2019 (COVID-19) Inpatients. *Arch. Pathol. Lab. Med.* **2022**, *146*, 34–37. [CrossRef]
11. Philippe, A.; Chocron, R.; Gendron, N.; Bory, O.; Beauvais, A.; Peron, N.; Khider, L.; Guerin, C.L.; Goudot, G.; Levasseur, F.; et al. Circulating Von Willebrand factor and high molecular weight multimers as markers of endothelial injury predict COVID-19 in-hospital mortality. *Angiogenesis* **2021**, *24*, 505–517. [CrossRef]
12. Hu, B.; Huang, S.; Yin, L. The cytokine storm and COVID-19. *J. Med. Virol.* **2021**, *93*, 250–256. [CrossRef]
13. Naß, J.; Terglane, J.; Gerke, V. Weibel Palade Bodies: Unique Secretory Organelles of Endothelial Cells that Control Blood Vessel Homeostasis. *Front. Cell Dev. Biol.* **2021**, *9*, 813995. [CrossRef]
14. Karampini, E.; Bierings, R.; Voorberg, J. Orchestration of Primary Hemostasis by Platelet and Endothelial Lysosome-Related Organelles. *Arterioscler. Thromb. Vasc. Biol.* **2020**, *40*, 1441–1453. [CrossRef]
15. Kawecki, C.; Lenting, P.J.; Denis, C.V. von Willebrand factor and inflammation. *J. Thromb. Haemost.* **2017**, *15*, 1285–1294. [CrossRef]

16. Cossutta, M.; Darche, M.; Carpentier, G.; Houppe, C.; Ponzo, M.; Raineri, F.; Vallée, B.; Gilles, M.E.; Villain, D.; Picard, E.; et al. Weibel-Palade Bodies Orchestrate Pericytes During Angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* **2019**, *39*, 1843–1858. [CrossRef]

17. Wilson, L.A.; McKeown, L.; Tumova, S.; Li, J.; Beech, D.J. Expression of a long variant of CRACR2A that belongs to the Rab GTPase protein family in endothelial cells. *Biochem. Biophys. Res. Commun.* **2015**, *456*, 398–402. [CrossRef]

18. Miteva, K.T.; Pedicini, L.; Wilson, L.A.; Jayasinghe, I.; Slip, R.G.; Marszalek, K.; Gaunt, H.J.; Bartoli, F.; Deivasigamani, S.; Sobradillo, D.; et al. Rab46 integrates Ca^{2+} and histamine signaling to regulate selective cargo release from Weibel-Palade bodies. *J. Cell Biol.* **2019**, *218*, 2232–2246. [CrossRef]

19. Srikanth, S.; Kim, K.D.; Gao, Y.; Woo, J.S.; Ghosh, S.; Calmettes, G.; Paz, A.; Abramson, J.; Jiang, M.; Gwack, Y. A large Rab GTPase encoded by CRACR2A is a component of subsynaptic vesicles that transmit T cell activation signals. *Sci. Signal* **2016**, *9*, ra31. [CrossRef]

20. Wu, B.; Rice, L.; Shrimpton, J.; Lawless, D.; Walker, K.; Carter, C.; McKeown, L.; Anwar, R.; Doody, G.M.; Srikanth, S.; et al. Biallelic mutations in calcium release activated channel regulator 2A (CRACR2A) cause a primary immunodeficiency disorder. *Elife* **2021**, *10*, e72559. [CrossRef]

21. Srikanth, S.; Jung, H.J.; Kim, K.D.; Souda, P.; Whitelegge, J.; Gwack, Y. A novel EF-hand protein, CRACR2A, is a cytosolic Ca2+ sensor that stabilizes CRAC channels in T cells. *Nat. Cell Biol.* **2010**, *12*, 436–446. [CrossRef] [PubMed]

22. Woo, J.S.; Srikanth, S.; Kim, K.D.; Elsaesser, H.; Lu, J.; Pellegrini, M.; Brooks, D.G.; Sun, Z.; Gwack, Y. CRACR2A-Mediated TCR Signaling Promotes Local Effector Th1 and Th17 Responses. *J. Immunol.* **2018**, *201*, 1174–1185. [CrossRef] [PubMed]

23. Talla, A.; Vasaikar, S.V.; Lemos, M.P.; Moodie, Z.; Lee Pebworth, M.-P.; Henderson, K.E.; Cohen, K.W.; Czartoski, J.L.; Lai, L.; Suthar, M.S.; et al. Longitudinal immune dynamics of mild COVID-19 define signatures of recovery and persistence. *bioRxiv* **2021**. [CrossRef]

24. Patel, M.A.; Knauer, M.J.; Nicholson, M.; Daley, M.; Van Nynatten, L.R.; Cepinskas, G.; Fraser, D.D. Organ and cell-specific biomarkers of Long-COVID identified with targeted proteomics and machine learning. *Mol. Med.* **2023**, *29*, 26. [CrossRef] [PubMed]

25. Bycroft, C.; Freeman, C.; Petkova, D.; Band, G.; Elliott, L.T.; Sharp, K.; Motyer, A.; Vukcevic, D.; Delaneau, O.; O'Connell, J.; et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **2018**, *562*, 203–209. [CrossRef] [PubMed]

26. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **2015**, *4*, 7. [CrossRef]

27. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef] [PubMed]

28. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef]

29. Schneider, T.D.; Stephens, R.M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097–6100. [CrossRef]

30. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

31. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [CrossRef]

32. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J.Y.; Wang, L.; Lupyan, D.; Dahlgren, M.K.; Knight, J.L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296. [CrossRef] [PubMed]

33. Kosmicki, J.A.; Horowitz, J.E.; Banerjee, N.; Lanche, R.; Marcketta, A.; Maxwell, E.; Bai, X.; Sun, D.; Backman, J.D.; Sharma, D.; et al. A catalog of associations between rare coding variants and COVID-19 outcomes. *medRxiv.* **2021**. [CrossRef]

34. Horowitz, J.E.; Kosmicki, J.A.; Damask, A.; Sharma, D.; Roberts, G.H.L.; Justice, A.E.; Banerjee, N.; Coignet, M.V.; Yadav, A.; Leader, J.B.; et al. Genome-wide analysis provides genetic evidence that ACE2 influences COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* **2022**, *54*, 382–392. [CrossRef]

35. Lee, I.G.; Cason, S.E.; Alqassim, S.S.; Holzbaur, E.L.F.; Dominguez, R. A tunable LIC1-adaptor interaction modulates dynein activity in a cargo-specific manner. *Nat. Commun.* **2020**, *11*, 5695. [CrossRef]

36. Boyle, A.L.; Rabe, M.; Crone, N.S.A.; Rhys, G.G.; Soler, N.; Voskamp, P.; Pannu, N.S.; Kros, A. Selective coordination of three transition metal ions within a coiled-coil peptide scaffold. *Chem. Sci.* **2019**, *10*, 7456–7465. [CrossRef]

37. Gifford, J.L.; Walsh, M.P.; Vogel, H.J. Structures and metal-ion-binding properties of the Ca2+-binding helix-loop-helix EF-hand motifs. *Biochem. J.* **2007**, *405*, 199–221. [CrossRef]

38. Misceo, D.; Holmgren, A.; Louch, W.E.; Holme, P.A.; Mizobuchi, M.; Morales, R.J.; De Paula, A.M.; Stray-Pedersen, A.; Lyle, R.; Dalhus, B.; et al. A dominant STIM1 mutation causes Stormorken syndrome. *Hum. Mutat.* **2014**, *35*, 556–564. [CrossRef]

39. Noury, J.B.; Böhm, J.; Peche, G.A.; Guyant-Marechal, L.; Bedat-Millet, A.L.; Chiche, L.; Carlier, R.Y.; Malfatti, E.; Romero, N.B.; Stojkovic, T. Tubular aggregate myopathy with features of Stormorken disease due to a new STIM1 mutation. *Neuromuscul. Disord.* **2017**, *27*, 78–82. [CrossRef]

40. Furuyama, W.; Shifflett, K.; Pinski, A.N.; Griffin, A.J.; Feldmann, F.; Okumura, A.; Gourdine, T.; Jankeel, A.; Lovaglio, J.; Hanley, P.W.; et al. Rapid protection from COVID-19 in nonhuman primates vaccinated intramuscularly but not intranasally with a single dose of a recombinant vaccine. *bioRxiv* **2021**. [CrossRef]

41. Wu, B.; Ramaiah, A.; Garcia, G., Jr.; Gwack, Y.; Arumugaswami, V.; Srikanth, S. ORAI1 establishes resistance to SARS-CoV-2 infection by regulating tonic type I interferon signaling. *bioRxiv* **2021**. [CrossRef]

42. Srikanth, S.; Woo, J.S.; Wu, B.; El-Sherbiny, Y.M.; Leung, J.; Chupradit, K.; Rice, L.; Seo, G.J.; Calmettes, G.; Ramakrishna, C.; et al. The Ca(2+) sensor STIM1 regulates the type I interferon response by retaining the signaling adaptor STING at the endoplasmic reticulum. *Nat. Immunol.* **2019**, *20*, 152–162. [CrossRef]

43. Zhang, Q.; Bastard, P.; Liu, Z.; Le Pen, J.; Moncada-Velez, M.; Chen, J.; Ogishi, M.; Sabli, I.K.D.; Hodeib, S.; Korol, C.; et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **2020**, *370*, eabd4570. [CrossRef]

44. Matuozzo, D.; Talouarn, E.; Marchal, A.; Zhang, P.; Manry, J.; Seeleuthner, Y.; Zhang, Y.; Bolze, A.; Chaldebas, M.; Milisavljevic, B.; et al. Rare predicted loss-of-function variants of type I IFN immunity genes are associated with life-threatening COVID-19. *Genome Med.* **2023**, *15*, 22. [CrossRef]