Title: Structural epitope profiling identifies antibodies associated with critical COVID-19 and long COVID

Authors: Patrick K.A. Kearns[1,2]*, Mihaly Badonyi[1], Kim Lee[3], Olivia Fleming[1], Lukas Gerasimivicous[1], Sam Benton[3], Jacky Guy[4], Scott Neilson[5], Helen Wise[6], Sara Jenks[6], Kate Templeton[6], Christina Dold[7], Teresa Lambe[7], Alexander J Mentzer[8], Julian C Knight[8], Andrew Pollard[7], COMBAT[8], Madhvi Menon[9], Tracy Hussell[9], CIRCO[9], Alan Carson[2], Laura McWhirter[2], Rennos Fragkoudis[5], Susan Rosser[9], David Cavanagh[3], Joseph A. Marsh[1], Dirk A. Kleinjan[4], Nick Gilbert[1]*

Affiliations:

[1]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh; Western General Hospital. Crewe Road, Edinburgh EH4 2XU.

[2]Centre for Clinical Brain Sciences, University of Edinburgh; 49 Little France Crescent, Edinburgh, United Kingdom EH16 4SB.

[3]Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh; The King's Buildings, Ashworth Laboratories, Charlotte Auerbach Road, Edinburgh, EH9 3FL.

[4]Centre for Synthetic and Systems Biology, School of Biological Sciences, University of Edinburgh; CH Waddington Building, Kings Buildings, Mayfield Road, Edinburgh EH9 3JD.

[5]Edinburgh Genome Foundry, University of Edinburgh; Michael Swann Building, Max Born Crescent, Edinburgh, United Kingdom EH9 3BF.

[6]Department of Clinical Biochemistry, Royal Infirmary of Edinburgh; 51 Little France Cres, Old Dalkeith Rd, Edinburgh, United Kingdom EH16 4SA.

[7]Oxford Vaccine Group, Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford; Churchill Hospital Old Road, Headington, Oxford, United Kingdom OX3 7LE.

[8]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Roosevelt Drive, Oxford, United Kingdom OX3 7BN.

[9]Division of Infection, Immunity & Respiratory Medicine, School of Biological Sciences, University of Manchester; Oxford Rd, Manchester, United Kingdom M13 9PL.

*Corresponding authors. Email: P.Kearns@ed.ac.uk; Nick.Gilbert@ed.ac.uk.

Abstract: Antibodies can have beneficial, neutral, or harmful effects so resolving an antibody repertoire to its target epitopes may explain heterogeneity in susceptibility to infectious disease. However, the three-dimensional nature of antibody-epitope interactions limits discovery of important targets. We describe and experimentally validated a computational method and synthetic biology pipeline for identifying structurally stable and functionally important epitopes from the SARS-CoV-2 proteome. We identify patterns of antibodies associated with immunopathology, including a non-isotype switching IgM response to a membrane protein epitope strongly associated with severe COVID-19 (adjusted OR 72.14, 95% CI: 9.71 – 1300.15). We suggest the mechanism is T independent B cell activation and identify persistence

(> 1 year) of this response in individuals with long COVID particularly affected by fatigue and depression. These findings may have implications for the ongoing medical and public health response to the pandemic.

One-Sentence Summary: Using a novel B cell epitope discovery method we have identified antibody signatures strongly associated with SARS-CoV-2 immunopathology and suggest the membrane protein is a pathological T independent antigen.

Main Text:

Most vertebrate species acquire remarkably heterogeneous repertoires of antibody-secreting cells after encountering a pathogen (*1*). B cells are selected for the affinity of their membrane-bound immunoglobulin receptor (mIg) to bind surfaces (epitopes) on foreign materials (e.g. viral proteins), and high-affinity clones are activated, expanded and differentiate to secrete antibodies into blood and mucosa (*2*). Anti-viral immunoglobulin binding may be functionally useful, for example, neutralising a pathogen; neutral; or harmful, for example, by causing antibody dependent enhancement (ADE) or autoimmunity (*3*).

The heterogeneity in the repertoire and in the functional effects of specific antibodies can therefore determine individual susceptibility to infection or disease (*4*), but the affinity of antibodies for three-dimensional structures in their native form makes deconvoluting the polyclonal B cell response challenging (*5, 6*). This contrasts with T cells, which inherently recognize only processed peptides derived from proteins. Antibodies typically bind multiple epitopes within each pathogen protein, such that an assay against a whole conformationally intact protein (e.g. ELISA) is an aggregate of the response and a large number of epitopes can be recognized across pathogen proteins by a population of individuals (*7*). Profiling antibody-epitope binding is therefore labour intensive and does not scale easily for most of the repertoire.

So-called “linear” B cell epitopes, those in which a continuous peptide fragment (10-20 amino acids) can bind an antibody, can be screened for using high-throughput methods thanks to their biochemical simplicity (*8*). However, these are a minor subset (~5-10%) of the whole repertoire (*6*), and most are thought to reflect functionally unhelpful antibodies that bind only degraded rather than whole proteins exposed on viable pathogens (*5, 6*). An alternative to experimental linear epitope discovery methods are in silico tools which predict the position of discontinuous epitopes (targets of most of the antibody repertoire) within the structure of a protein (*9*). However, to date, none of these tools attempt to identify whether a minimally sufficient set of residues can recapitulate predicted epitopes.

To address this, we have developed a computationally efficient thermodynamic method for predicting which parts of a protein can form stable peptides (10-100 amino acids in length). The idea is to identify sequences that would be most likely to adopt similar 3D conformations when synthesised as peptides compared with their context within the full-length protein. Our approach aims to enrich the candidate peptides for epitopes which we hypothesize are good candidates to be functionally relevant and immunodominant: as they are exposed on whole proteins and likely to persist in debris. Experimental efficiency is improved by 1/ not synthesizing peptides that are disordered and 2/ reducing false positives by eliminating stable peptides that will not adopt their native conformation.

We have applied this method to the SARS-CoV-2 proteome and validated the approach by characterising the antibody repertoire in independent clinical cohorts recruited within the UK National Health Service. SARS-CoV-2 is a good candidate for the discovery of beneficial and harmful targets as infection severity is variable (*10*) and antibodies are known to be both a key correlate of protection from infection (*11*) and a cause of harm (*12, 13*). However, COVID-19 outcome heterogeneity remains incompletely explained and the virus-specific factors that trigger breakdowns in tolerance and resultant immunopathology for SARS, MERS, and SARS-CoV-2 but not for other coronaviruses are not known. In addition, ADE has been described for coronaviruses (including SARS, FIPV and MERS) (*14, 15*), but it is not yet known whether components of the SARS-CoV-2-specific antibody response cause or enhance human disease.

Results

Validation of a computationally efficient approach to predicting structurally-stable immunogenic epitopes

A 1000 amino acid long protein has 86,086 possible sub-peptides between 10 and 100 residues in length. Across the SARS-CoV-2 proteome there are 1,240,901 such peptides. As it is expensive to screen this many peptides experimentally, or computationally using complex energy functions, we took advantage of a property that is relatively simple to compute from protein structure and is directly related to the energy of protein folding: the solvent-accessible surface area (SASA). Previously, we have demonstrated the utility of SASA for prediction of protein stability, flexibility, and assembly pathways, and shown that it is competitive with much more computationally intensive structural modelling strategies (*16–18*), making it feasible here to assess a huge number of possible peptides.

Starting with all proteins from the SARS-CoV-2 proteome (Fig. 1A), we selected those regions where structural models were available, and then fragmented the structures into all possible 10-100 amino acid sub-peptides. For all 489,207 of these sub-peptides, we computed a metric we term ΔASA_r, defined as the difference in SASA between the free peptide and the peptide in the context of the full protein structure normalised by the SASA of the peptide. Peptides with low ΔASA_r make fewer contacts outside of the peptide region and are thus more likely to maintain in the free form a conformation like that in the native structure (Fig. 1B).

Importantly, this strategy was not developed to predict immunogenic regions of the proteins, per se. Although we expect of two antibody-binding peptides the more stable peptide to be a better immunogen, all else being equal. Instead, the ‘stable’ peptides are most likely to adopt structural conformations as isolated peptides that are similar to the conformations in the full protein, and thus most likely to bind the same antibodies when expressed as peptides. Therefore, we complemented our structural method with other in silico predictors of immunogenicity (see Supplemental Methods), to identify stable peptides from regions most likely to be immunogenic. In total, we selected 100 peptides from the SARS-CoV-2 structural proteins (Fig. 1C), selecting those with low ΔASAr values, and seeking to obtain broad coverage of all proteins and a range of peptide sizes. The extra-virion exposed surfaces on the spike and membrane protein were covered in greatest depth, as these were pre-judged to be particularly likely to be functionally relevant. The nucleoprotein was also covered in depth as it is known to be immunodominant in and has been described as a cause of ADE in SARS (*19, 20*). We also selected 96 peptides from the non-structural proteins (Fig. 1D), again seeking broad coverage, but without prioritisation using ΔASA_r values.

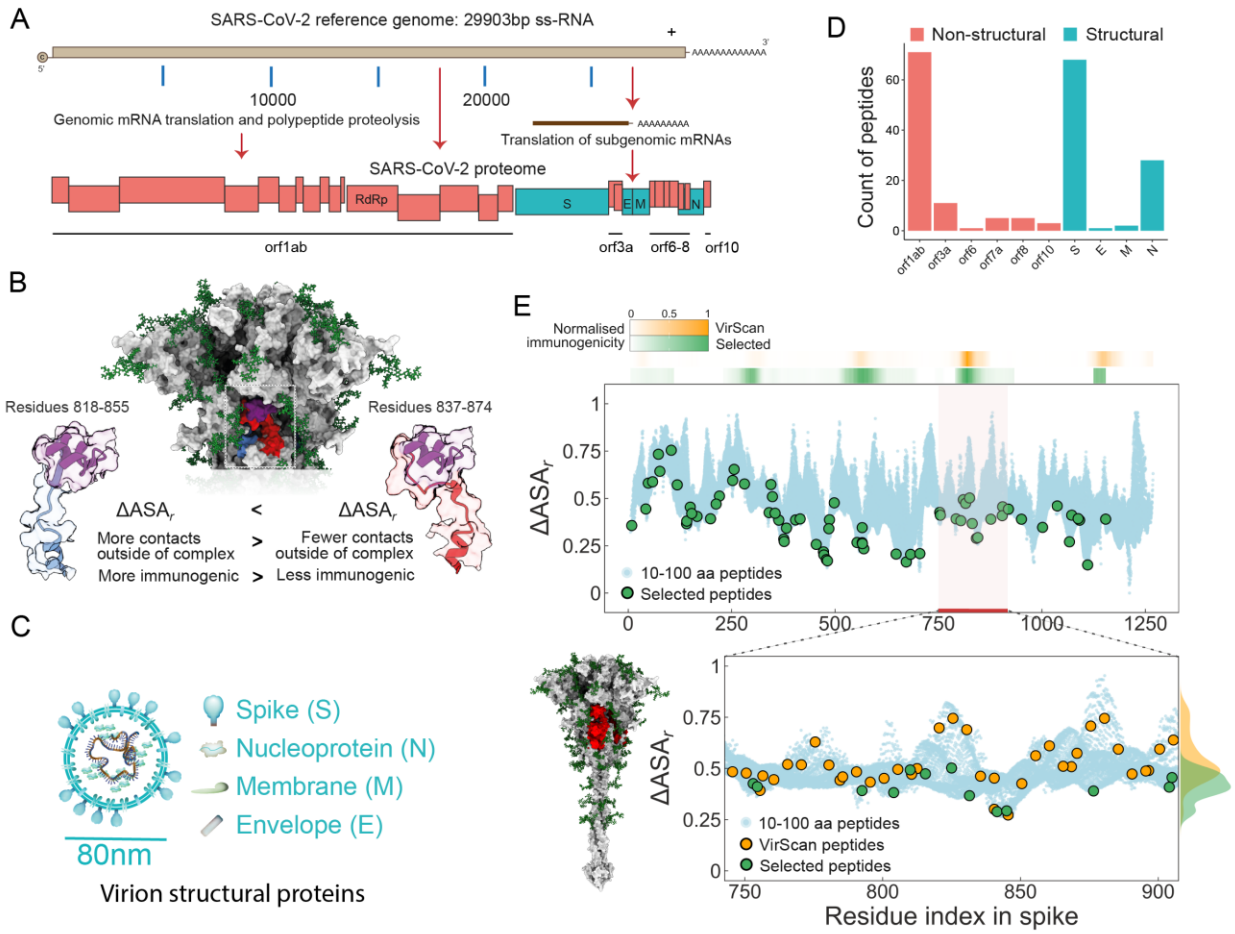

Fig. 1. Predicting structurally stable epitopes from the SARS-CoV-2 proteome. (A) SARS-CoV-2 reference genome and proteome. (Red – non-structural proteins; Cyan – structural proteins). (B) Peptides with lower ΔASA_r are more likely to adopt similar conformations as free peptides and are hypothesized to be more immunogenic than peptides covering the same region having higher ΔASA_r. (C) The structural proteins of a SARS-CoV-2 virion. (D) Count of the initial prioritised 196 peptides from SARS-CoV-2 proteome, by viral protein. (E) Top: the blue dots show the values of ΔASA_r for all possible 10-100 residue long peptides in spike along the linear sequence. Peptides are represented by their midpoints. Green dots mark the selected peptides using our structure-guided approach. Bars on top are co-linear with the residue index and show the immunogenic profile of spike as determined by VirScan phage-display (orange) (*21*) or by our approach (green). VirScan: Z-score difference, ours: ELISA ratio of positive sera to negative sera smoothened with a sliding window average of +/- 10aa and normalized to a scale of 0-1. Bottom: the amino acid region 750-900 of spike is shown in the structural context (left) or as a ΔASA_r distribution (right). Orange dots are midpoints of VirScan peptides, which generally have higher ΔASA_r values, as highlighted by the density diagram on the right.

Novel expression vectors for bacterial and mammalian expression and robot-assisted cloning and recombinant protein expression and purification pipelines were designed and optimised. After optimisation, reliable fusion protein expression and robotic purification by affinity chromatography was achieved. Isotype specific reactivities in sera were assayed for single epitopes using indirect colorimetric ELISAs (Fig. S1).

Commercially available (3M; Technopath, Tipperary, Ireland) pooled sera from SARS-CoV-2 recovered and pre-2019 naïve subjects was used to validate the reactivity of the predicted peptides (Fig. S2). We identified reactivity across structural and non-structural proteins of SARS-CoV-2 and compared these results to published studies using other methodologies (21, 22). The method broadly agreed with epitopes identified using other approaches but also identified reactivity where other techniques had not (Fig. 1E and Fig. S3). These analyses highlighted that the most immunogenic peptide amongst those which overlapped were often non-obvious to structurally agnostic approaches: it is not the case that shorter or longer peptides necessarily improved biophysical stability or immunogenicity (Fig.1E and Fig. S3).

To prioritise epitopes for further characterisation with scarce individual patient samples, we ranked peptides on immunogenicity by the ratio of reactivity in pooled positive sera to pooled negative sera (Fig. 2A). Figure 2A demonstrates findings for the structural protein antigens – the spike (S), nucleoprotein (N), membrane (M), and envelope (E). Consistent with other studies (*21*, *22*), we found that the most immunogenic epitopes were identified within the structural proteins, particularly S and N (Fig.2B), and a single external N-terminus epitope of the membrane protein (Fig.2C). After considering the positions of epitopes on the whole proteins and the immunogenicity ratios for IgG, IgA, and IgM, a subset of immunodominant non-overlapping peptides were selected for further characterisation (Fig.2A-C (red) and Fig. S2). The ratio of isotype-specific immunogenicity was similar for IgM:IgG, IgM:IgA,and IgA:IgG for all the epitopes with one exception. The N-terminus membrane protein antigen (M1) showed a markedly higher IgM versus other antibody isotypes compared to all the other reactive peptides (Fig.2D).

Examination of IgG responses in a reference set (NIBSC) of 23 individuals with PCR-confirmed SARS-CoV-2 and 14 individuals who were SARS-CoV-2 naïve confirmed the immunogenicity of the peptides at the individual level. In contrast to the (aggregate) results for whole viral proteins assayed with these sera (Fig. S4), we observed substantial inter-individual heterogeneity in the response resolved to level of epitopes (Fig.2E). Nonetheless, individual and combinations of the selected epitopes were able to discriminate with high sensitivity and specificity between infected and naïve individuals (ROC for combination of IgG responses AUC = 0.99, Fig. S5)

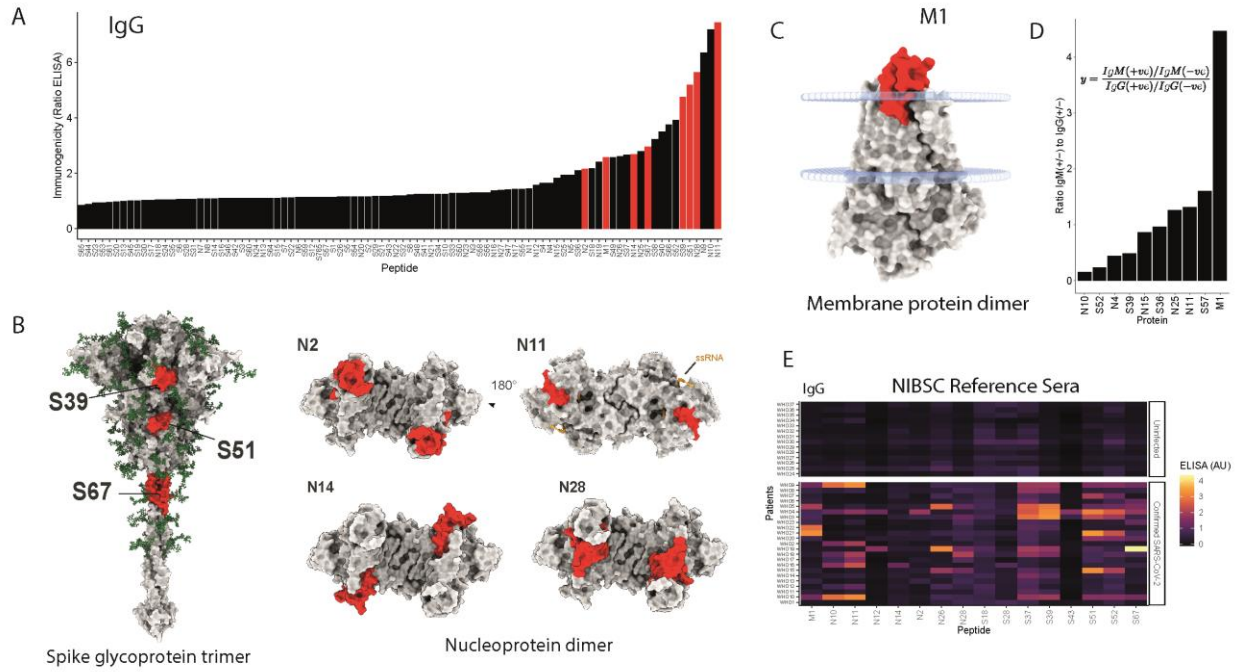

Fig. 2. Experimental validation of immunogenicity. (**A**) Selection (red) of eight prioritised non-overlapping peptides for further characterisation. Immunogenicity in bacterially expressed proteins (Y axis) as the ratios of the mean of at least three technical replicates for pooled positive sera to pooled negative sera (representative of at least two biological repeats (i.e. antigens expressed and purified independently) against individual peptides from the SARS-CoV-2 structural proteins (X axis). (B) Position of the selected (red) epitopes on the spike trimer and nucleoprotein dimer models. (C) Position of the M1 epitope on the extravirion surface of the membrane protein. (D) M1 epitope shows an unusual IgM dominant response. Ratio of IgM to IgG responses in pooled sera. Top 10 ratio values shown for all peptides with IgM and IgG > than lower limit of detection. (E) Heatmap of individual NIBSC reference sera (rows) against peptides (columns) for a selection of epitopes. NIBSC reference panel uninfected individuals (top panel) and individuals after confirmed SARS-CoV-2 infection (bottom panel). Demonstrating individual heterogeneity in the immune response to the eight selected peptides and other peptides from across the structural proteins of the virus.

Clinical cohorts reveal that antibody responses to these epitopes are common and epitopes are conserved through ongoing viral evolution.

We next characterised the reactivity to the eight prioritised antigens (Fig.2B&C) in independent clinical cohorts recruited from three UK centres (Edinburgh, Manchester, and Oxford). Individuals in these cohorts were infected with SARS-CoV-2 in the first pandemic wave in the UK: March-May 2020 (subject characteristics and inclusion/grading criteria in tables S1-5). Defining reactivity cut-offs as the mean response + 3SDs in the NIBSC negative controls, the proportions of individuals with detectable IgG responses in these cohorts were identified and found to be broadly similar across the cohorts despite differences in the proportion of individuals in each cohort who had required hospitalisation for COVID-19 (Fig. S6). The proportion of

individuals with IgG reactivity to at least one of the three prioritised spike proteins was >80% in each cohort and >90% of individuals had reactivity to at least one of the eight epitopes across S,N, and M proteins (Fig. S6).

Given viral antigenic evolution has led to the emergence of SARS-CoV-2 variants with substantial immune escape, we looked to see whether mutations had arisen in SARS-CoV-2 variants in these epitopes. Surprisingly, we found that none of the variants of interest or concern have acquired mutations in the prioritised S or N epitopes and only the recent Omicron variants have acquired mutations within the M1 epitope, including M:Q19E common to all omicron subvariants and subvariant specific mutations that may impact glycosylation at the D3 position (Fig. S6). The M protein M:Q19E mutation has been shown to be associated with increased fitness (*23*) and is notable in that the pH neutral glutamine at this location found in pre-Omicron SARS-CoV-2, is shared in SARS and SARS-related bat coronaviruses, whereas acidic amino acids (glutamic acid or aspartic acid) are observed in most other animal coronaviruses (Fig. S6 & Fig. S7).

M1 antibody isotype kinetics are atypical and IgM titres are predictive of known correlates of protection and severity.

Given the unusually high ratio of M1 IgM to other isotypes in pooled sera, we looked at IgM responses across a second independent negative cohort of 30 European individuals collected prior to 2019 to verify responses as genuine anti-SARS-CoV-2 reactivity. Consistent with findings from pooled negative sera, and the NIBSC reference sera, no pre-Covid cross-reactivity was observed in these samples for this epitope (Fig.3A). A high proportion of subjects in all three post-COVID clinical cohorts had IgM responses to the M1 antigen (Fig.3B).

Comparison of the Edinburgh cohort (of whom few were hospitalised) to the Manchester cohort (where all the individuals were hospitalised, required supplemental oxygen, and most had bilateral chest radiograph opacification), showed a significant tendency for higher M1 IgM titre in the hospitalised cohort (Fig. 3C). However, there are differences in the timing of recruitment and ages of subjects between these cohorts (table S2 and S3). Most subjects in both cohorts had detectable M1 IgM responses, and most lacked IgG responses (Fig.3C&D).

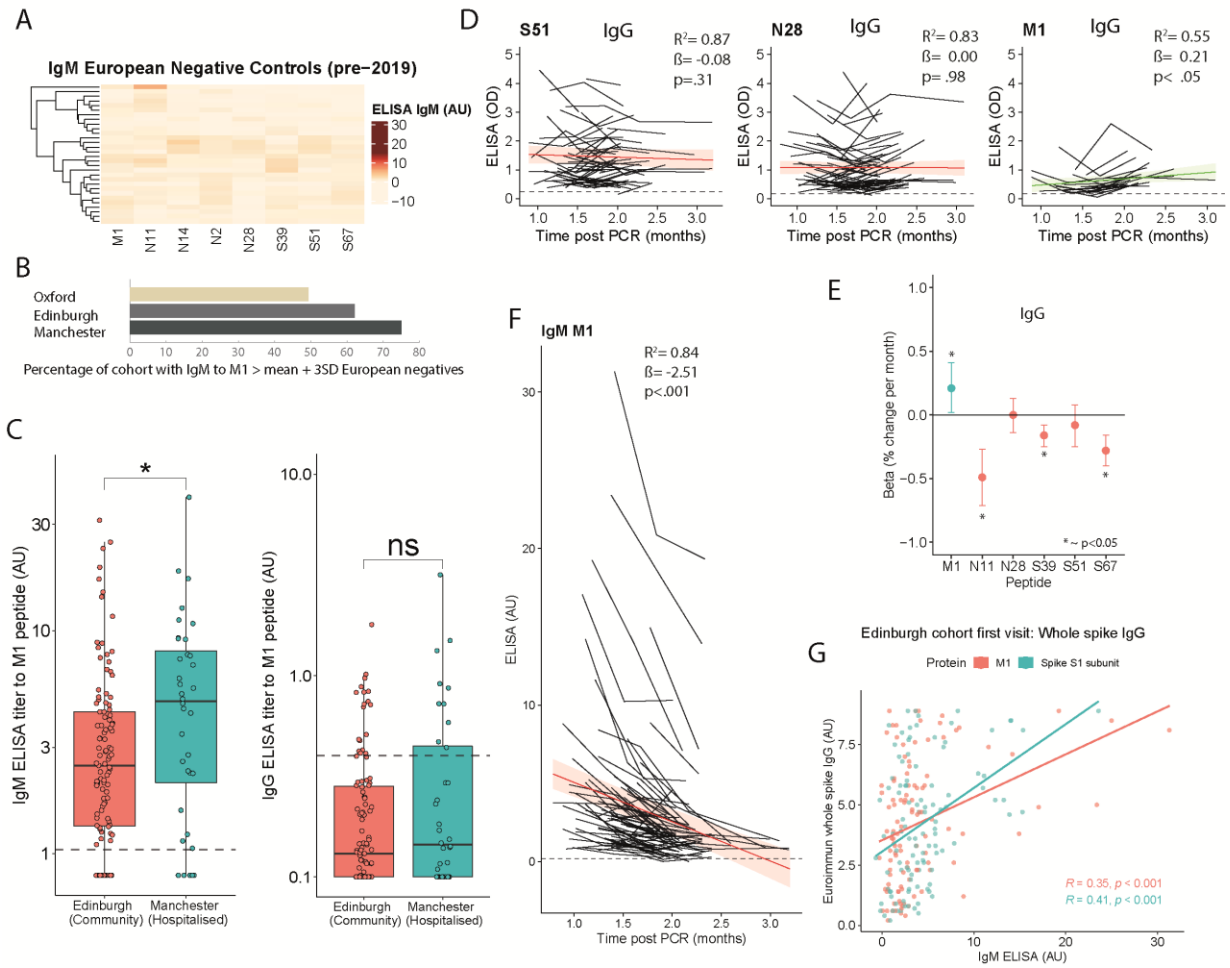

Fig. 3. The antibody kinetics for M1 antigen are unusual: high IgM responses with limited isotype class switching. (**A**) Heatmap of IgM response by ELISA for 30 European individuals collected pre-2019. (B) Proportion of three NHS clinical cohorts with M1 IgM above the mean + 3SD of the responses in A. (C) IgM and IgG titre to M1 in two clinical cohorts (Edinburgh and Manchester). Dashed line mean + 2SDs of responses for European negative controls (as in A). (D) Longitudinal samples IgG for Edinburgh cohort. Shown are individuals who have at least one positive result across study visits. Mixed effects random intercept models are fitted to determine the trends in trajectory. (E) Coefficients for time post PCR of random intercept models by antigen. M1 is the only epitope to show a significantly increasing titre in IgG over the three months post PCR. (F) IgM titres to M1 fall over the 3 months post infection and fall fastest for those with highest titres. (G) IgM to M1 and spike S1' subunit predict aggregate whole spike IgG titre measured by a Euroimmun assay.

To determine the antibody kinetics in the convalescent period, we fitted random intercept mixed effects models for isotype specific responses to individual epitopes for longitudinal follow up samples for the Edinburgh cohort (Fig. 3D). These revealed that the IgG response to most of the epitopes was static (N28, S51) or waned significantly (N11, S39, S67) over the 1-3 months post infection. The M1 epitope was again an exception, however (Fig. 3E). In those individuals where

M1 IgG responses could be detected, they increased significantly over the 3 months from low or undetectable initial levels (Fig. 3D&E). In contrast to IgG, the IgM response to the M1 antigen waned quickly from high levels in the 2.5 months post infection for most individuals, further consistent with a specific IgM dominated response having been provoked by SARS-CoV-2 infection (Fig. 3F).

The Manchester cohort were recruited earliest in the course of their infection (table S1), and in contrast to cohorts recruited later in the course of infection (tables S3 and S5), M1 IgM in these subjects showed a significant inverse association with age: i.e., the secreted IgM appears to be delayed in older persons requiring hospitalization. Delayed IgM secretory responses were evident for other epitopes to a lesser degree (Fig. S8). Natural and adaptive secreted IgM are known to play important non-redundant early roles in constraining viral infections and reduced IgM is a previously reported phenomenon associated with older age, high BMI, and male sex.

Surprisingly, M1 IgM predicted antibody responses to known correlates of protection and severe disease within the spike. M1 IgM predicted both whole spike IgG titre by a commercial assay (Fig. 3G) and pseudovirus neutralisation titre (Fig. S9). Unexpectedly, for both measures, M1 IgM was of similar predictive value as IgM to the whole S1' subunit of the spike itself. The Receptor Binding Domain (RBD) and N-terminus domain (NTD) are both regions within the S1' subunit of the spike (Fig. S6) and are known to be the site of most neutralising antibody binding (*24*) and consequently are important correlates of protection from infection (*25, 26*). However, these tend also to be higher in individuals following severe infection (*27*). This observation was replicated in an independent cohort, finding that M1 IgM predicted whole spike IgG in blood donors 2-6 months after infection, an M1 specific phenomenon. IgM titres for other peptides (including spike peptides) were not significantly associated with whole spike IgG (Fig. S9).

M1 position on the virion and antibody kinetics are suggestive of T independent B cell activation

The membrane protein is the most abundant SARS-CoV-2 protein and forms a homodimer that binds the three other virus structural proteins (S, E and N) (*28–32*). Each dimer is expected to present two extra-virion M1 epitopes (Fig. 4A). Structural analyses using cryo-electron microscopy and tomography by Neuman et al. report that dimers tend to be tightly positioned within coronavirus virions and that this organisation establishes virion shape with dimers repeating every 5^0 of radial curvature (*32*). These structural studies reveal that coronavirus membrane proteins are arranged in rhomboids with sides of approximately 4.5nm and 4nm and internal angle of 75^0, such that the distance between membrane protein dimers across the rhomboid is approximately 7.5nm in one axis and 3.8nm in the other (Fig. 4B) (*32*). The expression of the M, N, and E proteins is sufficient to form coronavirus like particles in cell culture (*28, 33*) and published analyses of virion structure and heterogeneity, show that approximately 10-20% of virions have no spike proteins and that the number of spike trimers per virion ranges between 0 and 50 with a mean of approximately 25 per virion (*34*) with virions tending to be pleomorphic in diameter and in the arrangement of spikes on their surfaces (Fig. 4B) (*34, 35*). In contrast, each virion presents approximately ~2,200 M1 epitopes (*32*). This suggests that some M1 specific B cells will encounter repetitive membrane bound M1 antigens on the surface of virions without obstruction by spike proteins.

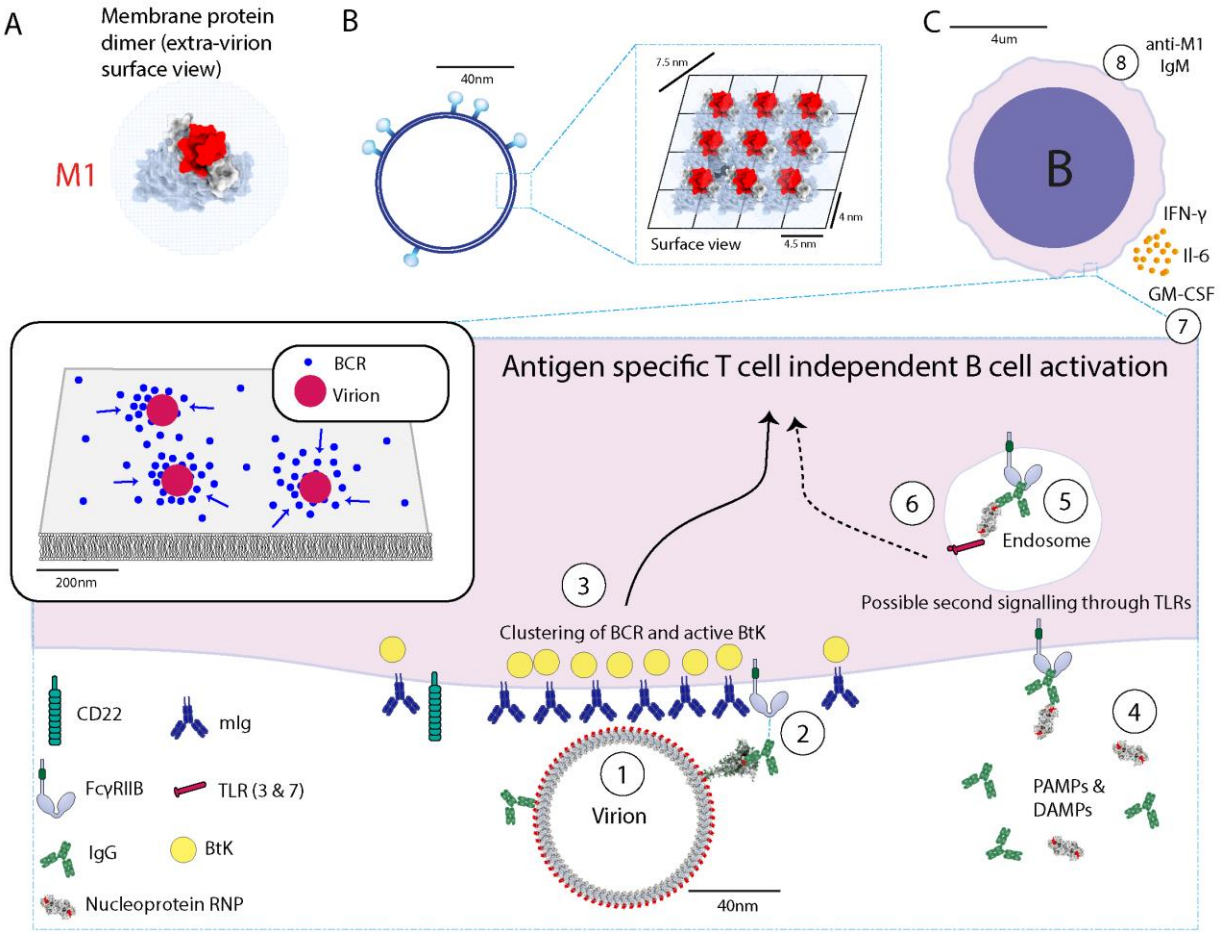

Fig. 4. Model of T independent B cell activation triggered by M1 antigen. (**A**) View of the membrane protein dimer from the extra-virion perspective. (B) Virions are pleomorphic in diameter and number of spikes with tightly arranged membrane proteins. This presents regular patterned extra-virion dimers, each expected to present two epitopes on the surface of the virus. (C) Model of T independent B cell activation roughly to scale. 1 – Virion arrangement of repeating M1 antigens, repeated approximately every 5^0 radially, engages with cognate membrane bound immunoglobulin (B cell receptors) on the surface of M1 specific B cells. 2 – secreted antibody (IgG/IgM) binding to surface epitopes can negatively regulate the TI response by competing with mIg binding or by triggering signaling through surface receptors on the B cell. 3 – Cross-linking activation of 10-20 clustered B cell receptors and active BtK is sufficient to trigger calcium influx resulting in B cell activation without T cell help. 4 – In the context of an active viral infection, the microenvironment is expected to include pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs), cytokines, and neutralizing and non-neutralising (e.g. anti-nucleoprotein) antibodies, 5 – B cell endosomes may take up various immunoglobulin-bound viral proteins, 6 - these may trigger regulatory signalling through toll like receptors, 7 – Cytokine release in response to TI activation, 8 – Antigen specific IgM secreted without isotype class switching.

Our suggested explanation for the atypical antibody kinetics of the M1 epitope is that the repeating organisation of the membrane protein leads to local activation of clustered B cell

receptors and clustered activation of Bruton's Tyrosine Kinase (BtK) on M1-specific B cells allowing for activation in the absence of follicular T cell help (Fig. 4C and Fig. S10). Antigen specific T independent (TI) B cell activation in this manner is typically associated with repetitive polysaccharide antigens positioned between 5 and 10nm apart and requires activation of a local cluster of a small number (10-20) of the ~10^5 total B cell receptors on a B cell membrane to generate the necessary signal strength (*36*). TI B cell activation is associated with extrafollicular plasmablast differentiation and the secretion of antigen specific IgM (37, 38). Isotype class switch recombination occurs predominantly in T dependent responses occurring in the T-cell:B-cell border of secondary lymphoid organs in germinal center destined B cells, where it is induced by helper T cell interaction and local cytokines (*36, 37, 39–43*). TI responses, in contrast, are characterised by limited isotype class switching as observed here for M1.

Secreted antibodies that bind to the repetitively presented epitopes (particularly IgG (*44, 45*) and IgM (*46*)) compete with mIg binding and are known to be important negative regulators of this mode of activation (*45–48*). Additional negative regulation involves signaling via inhibitory receptors that co-cluster with the B cell receptors including CD22 and Fc gamma receptors (*36*). Signaling via these receptors in response to their binding the Fc domain of IgG bound to the M1 epitope or other co-presented epitopes (most of which may be on the spike protein) may be important for negative regulation. This provides a mechanism by which non-neutralising IgG antibodies to exposed structural protein epitopes could have immune-regulatory effects.

The degree of TI B cell activation an antigen provokes is known to be dependent on several factors, including the antigen abundance (a function of stability and production), repetitiveness of the structure, binding kinetics, and binding competition from secreted antibodies (*36, 49–53*), but also well described is a role for signaling via toll like receptors regulating TI responses (*54–56*). In the context of a SARS-CoV-2 infection, it is plausible that toll like receptors in the endosomes of B cells recognize pathogen associated molecular patterns (PAMPs), for example, viral ssRNA-nucleoprotein-IgG complexes or other damage associated molecular patterns (DAMPs) (*54*). This offers a mechanism by which various anti-SARS-CoV-2 IgGs can regulate TI responses.

High M1 IgM is strongly associated with severe/critical COVID-19

TI responses are characterized by cytokine release from B cells and other co-activated immune cells including IFN-y, Il-6 and GM-CSF (*36*). These cytokines are also associated with severe COVID-19, with GM-CSF in particular associated with COVID-19 but not severe Influenza (*36, 57*). In addition, massive expansion of plasmablasts, absence of germinal centres, and high secreted IgM are hallmarks of critical and fatal cases of COVID-19 and may result from a dominant TI extrafollicular response (*58–60*). We therefore hypothesized that M1 specific TI may be an early event in acute SARS-CoV-2 immunopathology and that M1 IgM would therefore be associated with severe outcome as a marker of the degree of M1 TI activation.

To address this, we investigated the IgG, IgA, and IgM antibody responses to the eight viral epitopes and to the whole receptor binding domain of the spike protein (a gift from F. Krammer) in a cohort of individuals recruited in Oxford, UK in the first wave of the pandemic, who had either been asymptomatic/mild (non-hospitalised, n=45) or who had suffered severe/critical infection (n=25) (Fig. 5A). Using Lasso penalized logistic regression including the logarithm of participant age, sex and all the isotypes for each antigen and the whole RBD, the variables with the most predictive value were ascertained. The Lasso algorithm applies an increasingly strong

penalty shrinkage factor to multivariable regression coefficients such that they sequentially drop out of the model: initially improving model fit by biasing coefficients to null (the bias-variance trade off) allowing for variable selection. A high M1 IgM (defined as greater than the median value for the cohort) was the only antibody response that was a better predictor than the logarithm of age and was a stronger predictor of clinical severity than any other antibody response of any isotype including to the whole receptor binding domain (Fig. 5B).

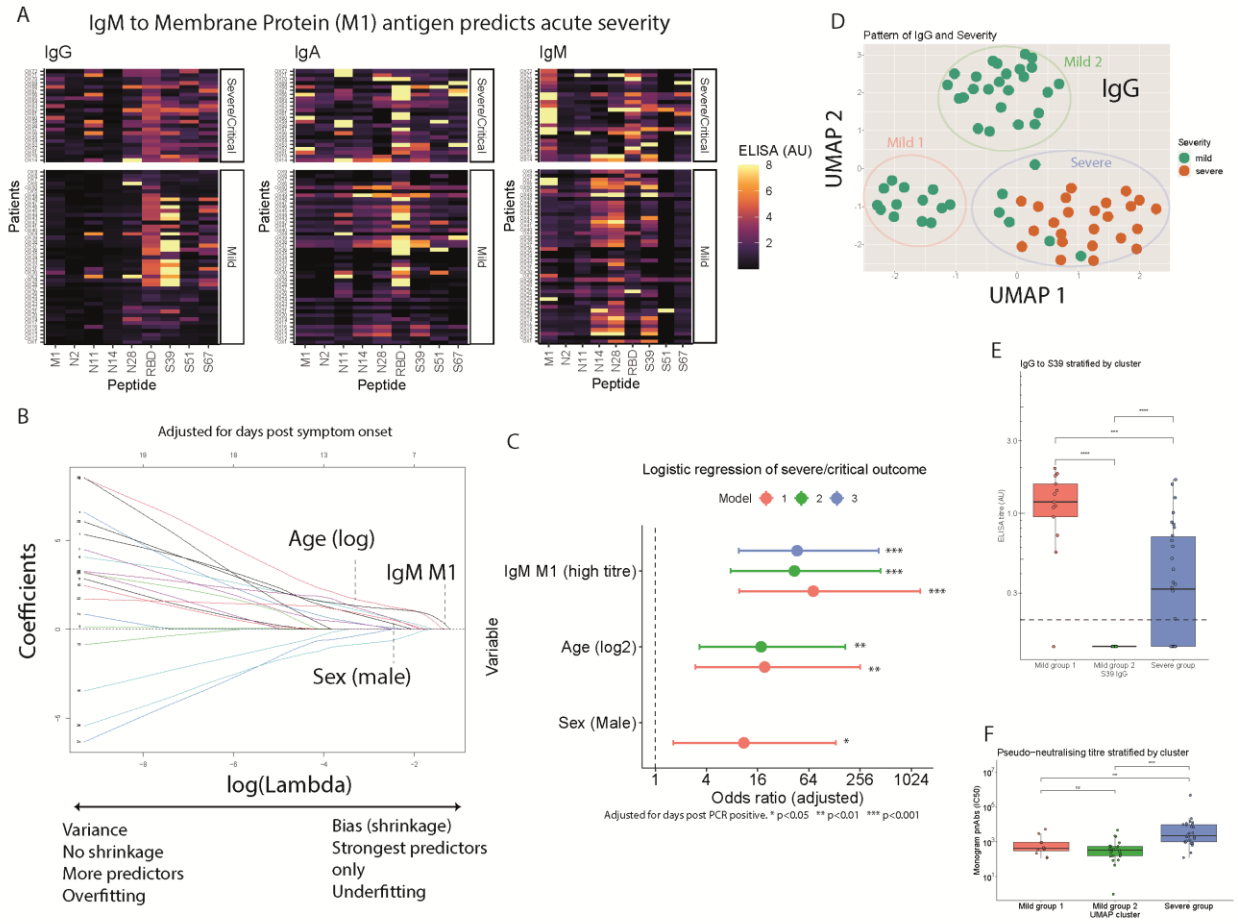

Fig. 5. M1 IgM is a strong predictor of severe/critical acute COVID-19. (**A**) Oxford cohort heatmap of individual sera (rows) against peptides (columns) for three immunoglobulin isotypes (column panels). Individuals represented by rows and separated into (row panels) by clinical severity of COVID-19 infection. (B) Penalized (Lasso) logistic regression demonstrates M1 IgM is the strongest predictor of the outcome (severe/critical COVID-19) of the ELISA responses to the eight peptides and receptor binding domain and age and sex. M1 IgM is the only antibody response that is retained in the model as a stronger predictor of severity than age. (C) Multi-variable logistic regression models (without penalisation) estimating the effect size and 95% confidence interval for the adjusted odds ratio. All models adjusted for days post PCR positive. (D) UMAP clustering of IgG responses in (A) with colouring based on severity. (E) S39 IgG by group assigned by cluster based on the patterns in D (red = mild 1; green = mild 2; blue = severe)

(F) Whole spike IgG titre (Jenner commercial assay) on Y-axis stratified by cluster pattern as in D (coloured as in E).

Subsequently three multivariable logistic regression models (without shrinkage) predicting severe/critical disease on M1 IgM were examined to estimate the unbiased effect size and relationship with other known predictors (age and male sex) adjusting for days post symptom onset (Fig. 5C and Table S6). These models demonstrated that having a high titre of M1 IgM was a strong predictor of severity, independent of the effects of age and sex. The point estimate on the effect size was strikingly large: in the model adjusting for age, sex and days post symptom onset, the adjusted odds ratio for a high M1 IgM was 72.14 (95% CI: 9.71 – 1300.15) whereas the odds ratio for doubling of age (a unit increase on log2 scale) was 19.33 (95% CI: 2.96 – 256.50). The size of this effect can be emphasized by considering that high M1 IgM was associated with an equivalent odds ratio as being 75 years old versus being 20 years old. In this cohort, the model including the coefficients for days since PCR; the demographic variables of age and sex; and the single immunological parameter of M1 IgM explained a high proportion of the variation in outcome (R2 (Tjur) = 0.65). For a 60-year-old man with a high M1 IgM versus a 30-year-old woman with a low M1 IgM, the point estimate for the combined OR was >15000, albeit with wide confidence intervals.

Patterns of IgG response are also associated with severe/critical COVID-19.

Uniform manifold approximation projection (UMAP) analyses of the isotype-specific responses identified clustering of IgG responses (Fig. 5D) but not the IgA and IgM responses (Fig. S10). The clustering algorithm (which is agnostic to the clinical severity and other characteristics of the participants), clearly grouped severe/critical participants together based on the pattern of their IgG responses and separated the mild group into two separate clusters. Clustering did not so clearly associate with other potential confounders like age, sex or time since infection (Fig. S11). The severe cohort was characterised by high IgG to N epitopes and one S epitope (S51), and low-medium IgG titres to the S39 epitope (Fig. 5A and Fig. S10).

The S39 epitope discriminated between the two mild cluster groups and showed a higher IgG titre in one cluster of mild patients than in the severe patients (Fig. 5F). This was notable as the severe group have higher anti-SARS-CoV-2 antibody titres overall (Fig. S10) and higher neutralising titres (Fig. 5F), consistent with other studies (26). The higher titres for S39, and to a lesser extent S67, in the mild group were IgG specific, a pattern that would be expected for any effects mediated through the Fc domain of binding IgG.

M1 IgM persists in a subset of convalescent individuals and is associated with long COVID and symptom burden.

Given M1 IgM titres appeared to wane relatively quickly in the months post infection (Fig. 3F), but that recovery time post-SARS-CoV-2 infection is variable, we were interested to determine whether there were subgroups of individuals for whom IgM responses persisted. To explore this, we tested IgM responses in a large cohort of convalescent individuals who had donated plasma to the Scottish National Blood Transfusion service in the 2-6 months after infection against the 8 peptide antigens. A subset of individuals (22/200) showed persistent M1 IgM (Fig. 6A). K-means (k=5) clustering identified three clusters of individuals with high or medium M1 IgM, one

large cluster with relatively low IgM to all the peptides, and a small cluster with IgM to S51 and N peptides (Fig. 6A).

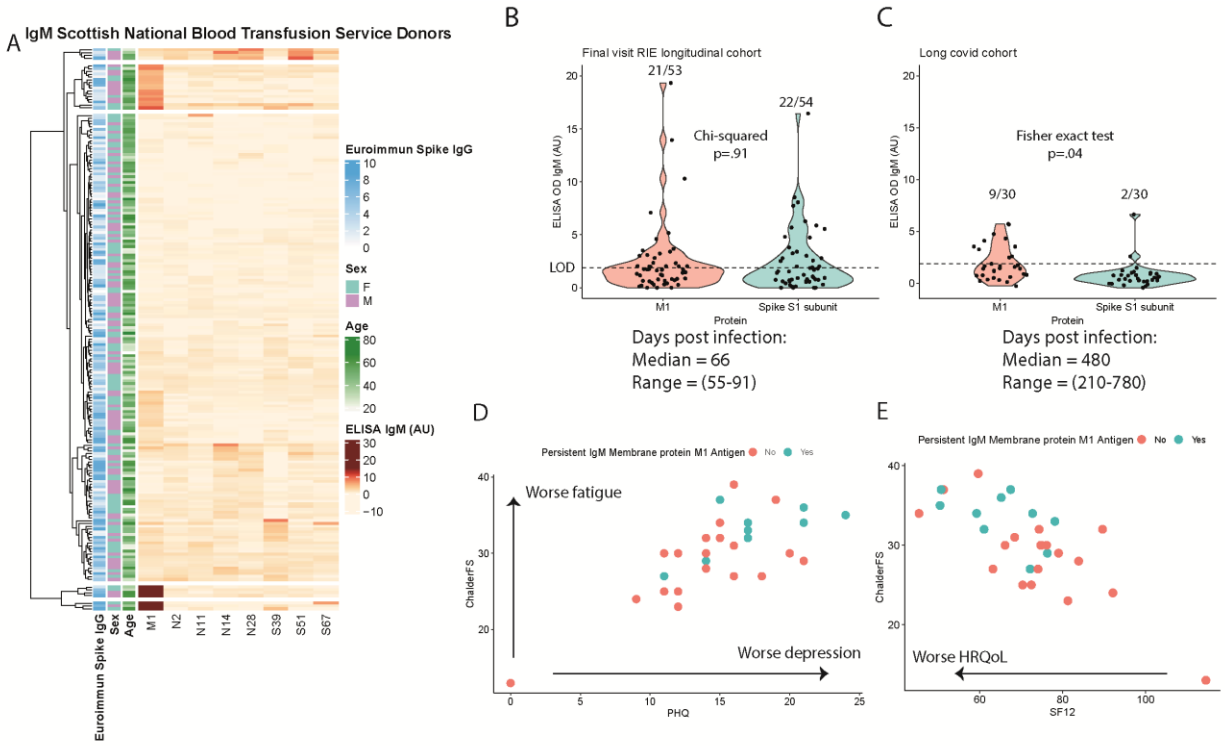

Fig. 6. Persistent M1 antibody responses are associated with long covid and long covid symptom burden. (**A**) Heatmap of Scottish National Blood Transfusion plasma donors who donated plasma after infection early in the coronavirus pandemic for trials of therapy with convalescent plasma. K-means clustering identifies two clusters with persistent very high IgM M1 (bottom and second bottom), one cluster with little persistent IgM (large middle group), one cluster with medium IgM M1 (second top), and one cluster with persistent S51 IgM and two nucleoproteins. (B) IgM responses in the final visit of the Edinburgh cohort to M1 antigen and a whole spike S1' subunit. The whole spike S1' subunit used rather than other epitopes because no single epitope is close to the M1 epitope for IgM publicness and spike S1' subunit likely reflects an aggregate of 100s of potential epitopes. LOD = limit of detection for calling positivity based on results in negative control subjects. (C) IgM responses in the long covid cohort to M1 antigen and a whole spike S1 subunit. (D) Chalder fatigue scale (y axis) against PHQ score for anxiety and depression. Subjects represented by points with persistent M1 IgM (cyan) and undetectable M1 IgM (red). (E) Chalder fatigue scale (y axis) against SF-12 score for a health-related quality of life score. SF-12 is comprised of two sub scores with mean 50 reflecting physical and mental domains of quality of life. Here these domains have been combined. Coloured as in D.

We next compared IgM titres to the whole spike S1' subunit and the M1 peptide in the final visit of the Edinburgh longitudinal cohort and found that 21/53 individuals were still positive for M1 IgM at a median of 66 days post infection and an almost identical proportion (22/54) individuals were still positive for IgM to the multi-epitope containing spike S1' subunit (p=.91, Fig. 6B). However, in contrast, repeating this analysis for a cohort of 30 individuals who had been referred by a healthcare provider to a long COVID study due to persistent cognitive symptoms (> 3

months) post infection (table S7), we found that 9/30 individuals were persistently positive for M1 IgM, whereas only 2/30 individuals were positive for IgM to the S1' subunit (p=.04, Fig. 6B). The median time since infection was substantially longer in the long COVID cohort than in any of our other cohorts (480 days). Those with persistent M1 IgM ranged from 240-780 days post infection at the time of sample draw.

In contrast to acute COVID-19, which is associated with a stereotypical clinical syndrome, long COVID is a less well-defined clinical entity where persistent symptoms are almost surely caused by a variety of mechanisms (*61, 62*). Amongst those who have persistent symptoms will be those with permanent organ damage from the acute illness (*61–66*), and individuals with non-specific symptoms due to other physical or mental illness with onset temporally coinciding with SARS-CoV-2 infection but mechanistically unrelated. Despite the expected heterogeneity, we found that even in this small cohort, persistent M1 IgM was associated with a significant 3.62 point worsening on the PHQ-15 score of anxiety and depression (p=0.048) and 4.53 point worsening on the Chalder fatigue scale (p=0.027) (Fig. 6C and Fig. S12). There was a non-significant tendency for a worse score (-8.47) on the SF-12 health related quality of life questionnaire which combines both physical and mental domains (p=.12) (Fig. S12).

Discussion

Numerous approaches exist for screening linear epitopes, however, library generation typically tiles k-mer peptides with arbitrary overlaps, so optimal biophysically stable epitope-containing peptides are usually not synthesized by chance (*8, 21*). Which of these k-mers adopt similar confirmations in isolation as on the whole protein scaffold is generally unknown. These structurally agnostic methods are of proven epidemiological value (*21, 67, 68*), but have low hit rates that limit downstream assays to high-throughput approaches which are not without limitations. Such systems include peptide arrays, or techniques that allow for physical linkage of the binding phenotype to genotype such bacteriophage or yeast where epitopes are displayed on structures more than 100-fold larger than the peptides themselves (*21, 67, 68*). Peptides that oligomerise these large display vehicles or aggregate with other peptides can be enriched or diminished, making quantitative interpretation challenging, which can be particularly an issue for pull-downs of multivalent antibody isotypes (e.g. divalent IgA or pentavalent IgM).

The method described here aims to greatly improve the efficiency of functionally important epitope discovery by incorporating information about the structural stability of the peptides at the outset. High yield predictions derived from experimentally confirmed or predicted protein structures can be validated experimentally in diverse downstream applications including low-, medium- or high-throughput techniques. We believe this method may have a wide range of possible applications, including in the design of peptide vaccines.

Using this approach and our synthetic biology fusion protein pipeline, we have identified numerous epitopes with clinically and immunologically important correlates from the SARS-CoV-2 proteome. In particular, a persistent IgM-with-limited-IgG/IgA response to an epitope in the extra-virion N-terminus of the membrane protein is a strong correlate of COVID-19 acute immunopathology. Notably, published studies profiling the antibody response to SARS-CoV-2 at the epitope level have tended to either focus on S or N, or have omitted testing IgM. Despite the vast literature on SARS-CoV-2 serology, to our knowledge only two studies have attempted to resolve the IgM response to specific epitopes on the SARS-CoV-2 membrane protein (*69, 70*) and no study has looked at clinical or immunological correlates of responses at epitope

resolution. Consistent with our findings, Jörrißen et al. found a high proportion (71.9%) of 32 subjects with IgM responses to a 20-mer peptide which is similar to our predicted optimal M1 (19-mer), and found a high IgM:IgG relative to another membrane protein antigen at two time points after infection (*70*). Wang et al. used 15-mer and 25-mers in a peptide array and found strong IgM reactivity in only one of eight individuals for the 15-mer, which may be missing important epitope defining residues, and no reactivity to the 25-mer peptide which includes extra hydrophobic residues from the transmembrane domain which we predict reduce stability (*69*). These results highlight the limitations of arbitrarily-sized, structure-agnostic k-mer approaches. One study has investigated clinical outcome associations of IgM responses to whole proteins in a mammalian cell-based assay, where the membrane is presumed to be presented in a similar fashion as on virions: exposing only the M1 epitope. Consistent with our findings, membrane protein specific IgM was associated with severe disease, although that method is unable to resolve the spike binding to the epitope level and so the unusual isotype pattern and strength of the IgM responses of the membrane epitope relative to other single epitopes is less apparent (*71*).

The association described here for M1 IgM is stronger than age, is independent of both age and sex, and is amongst the strongest correlates of severe COVID-19 described to date. For example, Bastard et al, have described rare functional auto-antibodies that neutralise 10ng/ml of both INF-α2 and IFN-ω at a 1:10 plasma dilution as associated with an OR of 67 (95% CI: 4–1109) predicting severe outcome (*76*). Their findings have been replicated in numerous independent cohorts (*77–82*) and ours have not, and the assay and statistical approaches differ. However, it is notable that the effect sizes are similar: adjusted OR 72 (95% CI: 9-1300) for a high M1 IgM in our study. Independent attempts to replicate our findings in other cohorts and to examine whether associations exist with other important immunological predictors such as auto-antibodies would be beneficial. The patients described with severe COVID-19 and anti-interferon antibodies have tended not to have had previous severe viral infections and we suggest that M1 TI responses may contribute to the early cytokine environment and immune response that explain the SARS-CoV-2 specific breakdown of tolerance that occurs in severe disease, including in individuals with other pre-existing potentially autoreactive lymphocytes, such as those antagonising type 1 interferons.

Finally, we found that persistent IgM responses to the membrane protein antigen are observed in approximately a third of individuals in a small and heterogenous, but well-phenotyped long COVID cohort. Biomarkers for long COVID are urgently needed to help stratify patients and to objectively score the outcome of trials of therapy. We found that persistent M1 IgM is characteristic of 11% (22/200) of blood donors in the 2-6 months after infection when IgM seroreversion to other epitopes had already occurred. Long term persistence was also evident in 30% (9/30) of individuals with long COVID at a median of 420 days after infection. Persistence was significantly associated with fatigue and anxiety/depression, symptoms which place a large burden on quality of life (*61, 62, 65, 83*).

Persistence of SARS-CoV-2 N antigens has been described in gut biopsies and autopsy tissues months after resolution of acute symptoms suggesting that reservoirs of viral materials may persist after acute infection in some tissues (*84*). As the membrane protein is the most abundant viral protein and chaperones the other structural proteins, it is plausible that continued antigen presentation of M1 will occur in addition to the described S and N persistence (*28, 29, 32, 34*). The findings of Martin et al. suggest that M1 can be presented on cell membranes, in addition to on virions (*71*). Whilst long COVID is undoubtedly mechanistically heterogenous, fatigue is consistently described as the most common disabling symptom (*65, 83, 85*). The biology of

fatigue in autoimmune and post-viral conditions is poorly understood, despite the importance of this symptom to patients. Identifying an immunological correlate is therefore encouraging and may be worthy of further investigation in other long COVID cohorts and similar approaches interrogating the role of B cell activation and IgM may be warranted in other post-viral and autoimmune syndromes. However, in our opinion these preliminary results should be interpreted cautiously, even if they are replicated elsewhere. Many of the participants in the long COVID cohort had identifiable and treatable clinical syndromes contributing to their symptom burden (e.g. depression and migraine) and whilst our results could be consistent with an immunological perturbance – immunological mechanisms have been hypothesized to underlie the aetiology of many such syndromes – much more work will be necessary to determine the possible clinical significance.

The lack of mutations in the S and N peptides, despite extensive adaptive evolution and antigenic escape to date, would be consistent with either these epitopes not being under selective pressures or the virus paying a high fitness penalty for mutations in these structurally stable domains. Severe COVID-19 is a late event in the course of infection, whereas most transmission occurs early (*86*), and so a lack of selective pressure would not necessarily rule out a role for these epitopes determining disease severity. The exception to the conservation across variants is in M1 where mutations have arisen in the Omicron variants (Fig. S6) (*87*). The M protein of coronaviruses is a well-studied model of N- and O-linked glycosylation and strain-specific variation in M glycosylation is thought to determine organ-specific tropism and pathogenicity of murine hepatitis coronavirus by incompletely understood mechanisms (*88–91*). BA.5 has an D3N mutation which may be associated with novel N-glycosylation, and there are now at least three polymorphisms circulating in the Omicron subvariants at this position. M1 mutations may be worth considering alongside the differences variant spikes when investigating intrinsic differences in disease phenotype or immunogenicity.

The observational nature of our study means that the proposed mechanism for the M1 antibody profile and associations is necessarily speculative. However, we consider it an attractive hypothesis for further testing for the following reasons: 1/ The cytokine and immunological response associated with TI antigens is known to overlap the biomarkers of COVID-19 specific immunopathology observed in large cohort studies (particularly high IFN-gamma, GM-CSF, soluble IgM and plasmablasts numbers) (*58–60*); 2/ structural analyses of the membrane protein within the virion suggest that local B cell receptor (BCR) clustering is physically plausible (*32*); 3/ A TI response explains a high IgM response with limited isotype class switching described here consistent with a previous report (*36*); 4/ delayed soluble IgM associated with older age may contribute to the age-severity association in Covid-19 due to less negative feedback of TI activation; 5/ young children do not mount TI antigen-specific responses potentially contributing to their relative sparing from acute COVID-19 immunopathology (*36*); 6/ TI activation of B cells may explain why SARS-CoV-2, more than other viruses is associated with breakdowns in tolerance, and tissue-specific and temporal uncoupling of inflammation from viral load (*12, 13, 92*); 7/ modulation of TI B cell activation by toll like receptor signaling within B cell endosomes is a plausible route to explain anti-N coronavirus IgG-mediated antibody dependent enhancement described for SARS and rare monogenic susceptibilities to severe COVID-19 (*44–46, 54, 55*); 8/ co-clustering of negative regulators, like CD22 and FcγR with the BCR, provides an actionable mechanism by which non-neutralising IgG binding on the virion to conserved targets may downregulate immunopathology contributing to the decoupling of protection from infection and protection from severity observed after vaccination despite extensive viral antigenic evolution at the sites of neutralising antibody binding (*36, 45, 46*); and 9/ the N termini of SARS and SARS-

CoV-2 membrane proteins show high homology, suggesting that this mechanism may be relevant to other related viruses which cause immunopathology. A clear line of sight to therapeutic or preventative intervention would exist if the findings can be replicated and the mechanism confirmed.

References and Notes:

1. M. F. Flajnik, Comparative analyses of immunoglobulin genes: surprises and portents. Nat. Rev. Immunol. 2, 688–698 (2002).
2. T. Kurosaki, K. Kometani, W. Ise, Memory B cells. Nat. Rev. Immunol. 15, 149–59 (2015).
3. S. B. Halstead, Neutralization and Antibody-Dependent Enhancement of Dengue Viruses. Adv. Virus Res. 60, 421–467 (2003).
4. A. T. Huang et al., A systematic review of antibody mediated immunity to coronaviruses: kinetics, correlates of protection, and association with severity. Nat. Commun. 11, 1–16 (2020).
5. W. G. Laver et al., Epitopes on protein antigens: misconceptions and realities. Cell. 61, 553–556 (1990).
6. M. H. V. Van Regenmortel, What is a b-cell epitope? Methods Mol. Biol. 524, 3–20 (2009).
7. J. A. Schroer et al., Mapping epitopes on the insulin molecule using monoclonal antibodies. Eur. J. Immunol. 13, 693–700 (1983).
8. M. H. V. Van Regenmortel, Specificity, polyspecificity, and heterospecificity of antibody-antigen recognition. J. Mol. Recognit. 27, 627–639 (2014).
9. P. Haste Andersen, M. Nielsen, O. Lund, Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci. 15, 2558–2567 (2006).
10. M. O'Driscoll et al., Age-specific mortality and immunity patterns of SARS-CoV-2. Nature. 590, 140–145 (2021).
11. T. Zohar, G. Alter, Dissecting antibody-mediated protection against SARS-CoV-2. Nat. Rev. Immunol. 20 (2020), pp. 392–394.
12. P. Bastard et al., Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science (80-.). 370 (2020), doi:10.1126/science.abd4585.
13. E. Y. Wang et al., Diverse Functional Autoantibodies in Patients with COVID-19. Nature, 1–6 (2021).
14. M. Jaume et al., Anti-Severe Acute Respiratory Syndrome Coronavirus Spike Antibodies Trigger Infection of Human Immune Cells via a pH- and Cysteine Protease-Independent Fc R Pathway. J. Virol. 85, 10582–10597 (2011).
15. S. F. Wang et al., Antibody-dependent SARS coronavirus infection is mediated by antibodies against spike proteins. Biochem. Biophys. Res. Commun. 451, 208–214 (2014).

16. J. A. Marsh, Buried and accessible surface area control intrinsic protein flexibility. J. Mol. Biol. 425, 3250–3263 (2013).

17. J. A. Marsh, S. A. Teichmann, Relative solvent accessible surface area predicts protein conformational changes upon binding. Structure. 19, 859–867 (2011).

18. J. A. Marsh et al., Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. Cell. 153, 461 (2013).

19. B. Meyer, C. Drosten, M. A. Müller, Serological assays for emerging coronaviruses: challenges and pitfalls. Virus Res. 194, 175–183 (2014).

20. F. Yasui et al., Prior immunization with severe acute respiratory syndrome (SARS)-associated coronavirus (SARS-CoV) nucleocapsid protein causes severe pneumonia in mice infected with SARS-CoV. J. Immunol. 181, 6337–6348 (2008).

21. E. Shrock et al., Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. Science (80-.). 370 (2020), doi:10.1126/science.abd4250.

22. C. R. Zamecnik et al., ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. Cell Reports Med. 1, 100123 (2020).

23. F. Obermeyer et al., Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. Science (80-.). (2022), doi:10.1126/SCIENCE.ABM1208.

24. L. Piccoli et al., Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. Cell. 183, 1024-1042.e21 (2020).

25. A. Addetia et al., Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. J. Clin. Microbiol. 58 (2020), doi:10.1128/JCM.02107-20.

26. S. A. Plotkin, Immunologic correlates of protection induced by vaccination. Pediatr. Infect. Dis. J. 20, 63–75 (2001).

27. E. H. Y. Lau et al., Neutralizing antibody titres in SARS-CoV-2 infections. Nat. Commun. 2021 121. 12, 1–7 (2021).

28. B. Boson et al., The SARS-CoV-2 envelope and membrane proteins modulate maturation and retention of the spike protein, allowing assembly of virus-like particles. J. Biol. Chem. 296, 100111 (2021).

29. K. Narayanan et al., Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J. Virol. 74, 8127–8134 (2000).

30. C. A. M. de Haan, H. Vennema, P. J. M. Rottier, Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. J. Virol. 74, 4967 (2000).

31. Y. L. Siu et al., The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. J. Virol. 82, 11318–11330 (2008).

32. B. W. Neuman et al., A structural analysis of M protein in coronavirus assembly and morphology. J. Struct. Biol. 174, 11–22 (2011).

33. Y. L. Siu et al., The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. J. Virol. 82, 11318–11330 (2008).

34. Z. Ke et al., Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nat. 2020 5887838. 588, 498–502 (2020).

35. L. Tai et al., Nanometer-resolution in situ structure of the SARS-CoV-2 postfusion spike protein. Proc. Natl. Acad. Sci. U. S. A. 118 (2021), doi:10.1073/PNAS.2112703118/SUPPL_FILE/PNAS.2112703118.SM02.MP4.

36. J. J. Mond et al., T cell independent antigens. Curr. Opin. Immunol. 7, 349–354 (1995).

37. T. A. Y. Shih, M. Roederer, M. C. Nussenzweig, Role of antigen receptor affinity in T cell-independent antibody responses in vivo. Nat. Immunol. 3, 399–406 (2002).

38. J. Stavnezer, J. E. J. Guikema, C. E. Schrader, Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. 26 (2008), pp. 261–292.

39. B. A. Heesters et al., Antigen Presentation to B Cells. Trends Immunol. 37, 844–854 (2016).

40. Y. R. Carrasco, F. D. Batista, B cell recognition of membrane-bound antigen: an exquisite way of sensing ligands. Curr. Opin. Immunol. 18, 286–291 (2006).

41. D. Paus et al., Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. J. Exp. Med. 203, 1081–1091 (2006).

42. C. García de Vinuesa et al., T-independent type 2 antigens induce B cell proliferation in multiple splenic sites, but exponential growth is confined to extrafollicular foci, doi:10.1002/(SICI)1521-4141(199904)29:04.

43. J. A. Roco et al., Class-Switch Recombination Occurs Infrequently in Germinal Centers. Immunity. 51, 337-350.e7 (2019).

44. T. V. Obukhanych, M. C. Nussenzweig, T-independent type II immune responses generate memory B cells. J. Exp. Med. 203, 305–310 (2006).

45. B. Heyman, Feedback regulation by IgG antibodies. Immunol. Lett. 88, 157–161 (2003).

46. M. Brüggemann, K. Rajewsky, Regulation of the antibody response against hapten-coupled erythrocytes by monoclonal antihapten antibodies of various isotypes. Cell. Immunol. 71, 365–373 (1982).

47. T. Takai et al., Augmented humoral and anaphylactic responses in Fc gamma RII-deficient mice. Nature. 379, 346–349 (1996).

48. P. H. Brodeur, H. H. Wortis, Regulation of thymus-independent responses: unresponsiveness to a second challenge of TNP-Ficoll is mediated by hapten-specific antibodies. J. Immunol. 125 (1980).

49. R. A. Herlands et al., T cell-independent and toll-like receptor-dependent antigen-driven activation of autoreactive B cells. Immunity. 29, 249–260 (2008).

50. T. Fehr et al., T cell-independent type I antibody response against B cell epitopes expressed repetitively on recombinant virus particles. Proc. Natl. Acad. Sci. U. S. A. 95, 9477–9481 (1998).

51. D. R. Milich, A. Mclachlan, The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen. Science. 234, 1398–1401 (1986).

52. D. Paus et al., Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. J. Exp. Med. 203, 1081–1091 (2006).

53. K. M. Toellner et al., Immunoglobulin switch transcript production in vivo related to the site and time of antigen-specific B cell activation. J. Exp. Med. 183, 2303–2312 (1996).

54. E. A. Leadbetter et al., Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. Nature. 416, 603–607 (2002).

55. A. Lanzavecchia, F. Sallusto, Toll-like receptors and innate immunity in B-cell activation and antibody responses. Curr. Opin. Immunol. 19, 268–274 (2007).

56. E. J. Pone et al., BCR-signalling synergizes with TLR-signalling for induction of AID and immunoglobulin class-switching through the non-canonical NF-κB pathway. Nat. Commun. 3, 767 (2012).

57. R. S. Thwaites et al., Inflammatory profiles across the spectrum of disease reveal a distinct role for GM-CSF in severe COVID-19. 6 (2021) (available at https://www.science.org/doi/full/10.1126/sciimmunol.abg9873).

58. N. Kaneko et al., Loss of Bcl-6-Expressing T Follicular Helper Cells and Germinal Centers in COVID-19. Cell. 183, 143-157.e13 (2020).

59. S. De Biasi et al., Expansion of plasmablasts and loss of memory B cells in peripheral blood from COVID-19 patients with pneumonia. Eur. J. Immunol. 50, 1283–1294 (2020).

60. J. P. Bernardes et al., Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. Immunity. 53, 1296-1314.e9 (2020).

61. D. L. Sykes et al., Post-COVID-19 Symptom Burden: What is Long-COVID and How Should We Manage It? Lung. 199, 113–119 (2021).

62. M. Whitaker et al., Persistent COVID-19 symptoms in a community study of 606,434 people in England. Nat. Commun. 13 (2022), doi:10.1038/S41467-022-29521-Z.

63. A. J. Morrow et al., A multisystem, cardio-renal investigation of post-COVID-19 illness. Nat. Med. 2022, 1–11 (2022).

64. R. J. J. van Gassel et al., High prevalence of pulmonary sequelae at 3 months after hospital discharge in mechanically ventilated survivors of COVID-19. Am. J. Respir. Crit. Care Med. 203, 371–374 (2021).

65. Z. Al-Aly, Y. Xie, B. Bowe, High-dimensional characterization of post-acute sequelae of COVID-19. Nature. 594, 259–264 (2021).

66. Y. Xie et al., Long-term cardiovascular outcomes of COVID-19. Nat. Med. 2022 283. 28, 583–590 (2022).

67. R. D. Schubert et al., Pan-viral serology implicates enteroviruses in acute flaccid myelitis. Nat. Med. 25, 1748 (2019).

68. M. J. Mina et al., Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. Science. 366, 599–606 (2019).

69. H. Wang et al., SARS-CoV-2 Proteome Microarray for Mapping COVID-19 Antibody Interactions at Amino Acid Resolution. ACS Cent. Sci. 6, 2238–2249 (2020).

70. P. Jörrißen et al., Antibody Response to SARS-CoV-2 Membrane Protein in Patients of the Acute and Convalescent Phase of COVID-19. Front. Immunol. 12, 679841 (2021).

71. S. Martin et al., SARS-CoV-2 integral membrane proteins shape the serological responses of patients with COVID-19. iScience. 24, 103185 (2021).

72. M. F. Bachmann, H. Hengartner, R. M. Zinkernagel, T helper cell-independent neutralizing B cell response against vesicular stomatitis virus: role of antigen patterns in B cell induction? Eur. J. Immunol. 25, 3445–3451 (1995).

73. E. Szomolanyi-Tsuda et al., Antiviral T-cell-independent type 2 antibody responses induced in vivo in the absence of T and NK cells. Virology. 280, 160–168 (2001).

74. D. Depoil et al., CD19 is essential for B cell activation by promoting B cell receptor-antigen microcluster formation in response to membrane-bound ligand. Nat. Immunol. 9, 63–72 (2008).

75. R. A. Crowther et al., Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. Cell. 77, 943–950 (1994).

76. P. Bastard et al., Autoantibodies neutralizing type I IFNs are present in ~4% of uninfected individuals over 70 years old and account for ~20% of COVID-19 deaths. Sci. Immunol. 6 (2021), doi:10.1126/sciimmunol.abl4340.

77. X. Solanich et al., Pre-existing Autoantibodies Neutralizing High Concentrations of Type I Interferons in Almost 10% of COVID-19 Patients Admitted to Intensive Care in Barcelona. J. Clin. Immunol. 41 (2021) (available at /pmc/articles/PMC8475877/).

78. J. Troya et al., Neutralizing Autoantibodies to Type I IFNs in >10% of Patients with Severe COVID-19 Pneumonia Hospitalized in Madrid, Spain. J. Clin. Immunol. 41, 914–922 (2021).

79. R. Koning et al., Autoantibodies against type I interferons are associated with multi-organ failure in COVID-19 patients. Intensive Care Med. 47, 704–706 (2021).

80. A. Chauvineau-Grenier et al., Autoantibodies Neutralizing Type I Interferons in 20% of COVID-19 Deaths in a French Hospital. J. Clin. Immunol. 42, 459–470 (2022).

81. M. G. P. van der Wijst et al., Type I interferon autoantibodies are associated with systemic immune alterations in patients with COVID-19. Sci. Transl. Med. 13, 29 (2021).

82. L. F. Chen, C. De Yang, X. B. Cheng, Anti-Interferon Autoantibodies in Adult-Onset Immunodeficiency Syndrome and Severe COVID-19 Infection. Front. Immunol. 12 (2021), doi:10.3389/FIMMU.2021.788368.

83. F. J. Carod-Artal, Post-COVID-19 syndrome: Epidemiology, diagnostic criteria and pathogenic mechanisms involved. Rev. Neurol. 72, 384–396 (2021).

84. C. Gaebler et al., Evolution of antibody immunity to SARS-CoV-2. Nat. 2021 5917851. 591, 639–644 (2021).

85. M. Taquet et al., Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. PLOS Med. 18, e1003773 (2021).

86. H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, Serial interval of novel coronavirus (COVID-19) infections. Int. J. Infect. Dis. (2020), doi:10.1016/j.ijid.2020.02.060.

87. S. Khare et al., GISAID's Role in Pandemic Response. China CDC Wkly. 3, 1049 (2021).

88. C. A. M. De Haan et al., The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain. Virology. 312, 395 (2003).

89. S. Alexander, J. H. Elder, Carbohydrate dramatically influences immune reactivity of antisera to viral glycoprotein antigens. Science. 226, 1328–1330 (1984).

90. K. V. Holmes, E. W. Doller, L. S. Sturman, Tunicamycin resistant glycosylation of coronavirus glycoprotein: demonstration of a novel type of viral glycoprotein. Virology. 115, 334–344 (1981).

91. M. Oostra et al., Glycosylation of the Severe Acute Respiratory Syndrome Coronavirus Triple-Spanning Membrane Proteins 3a and M. J. Virol. 80, 2326 (2006).

92. D. A. Dorward et al., Tissue-specific immunopathology in fatal COVID-19. Am. J. Respir. Crit. Care Med. 203, 192–201 (2021).

93. A. M. Waterhouse et al., Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 25, 1189–1191 (2009).

94. F. Muecksch et al., Longitudinal serological analysis and neutralizing antibody levels in coronavirus disease 2019 convalescent patients. J. Infect. Dis. 223, 389–398 (2021).

95. E. F. Pettersen et al., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82 (2021).

96. M. A. Lomize et al., OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res. 40, D370–D376 (2012).

97. L. A. Kelley et al., The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 2015 106. 10, 845–858 (2015).

98. L. Heo, H. Lee, C. Seok, GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. Sci. Rep. 6 (2016), doi:10.1038/SREP32153.

99. H. M. Berman et al., The Protein Data Bank. Nucleic Acids Res. (2000), , doi:10.1093/nar/28.1.235.

100. M. D. Winn et al., Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. 67 (2011), pp. 235–242.

101. S. Miller et al., Interior and surface of monomeric proteins. J. Mol. Biol. 196, 641–656 (1987).

102. M. C. Jespersen et al., BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res. 45, W24–W29 (2017).

103. C. Zhou et al., SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. Nucleic Acids Res. 47, W388–W394 (2019).

104. J. V. Kringelum et al., Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. PLoS Comput. Biol. 8 (2012), doi:10.1371/JOURNAL.PCBI.1002829.

105. M. Kurcinski et al., CABS-flex standalone: A simulation environment for fast modeling of protein flexibility. Bioinformatics. 35, 694–695 (2019).

106. J. Ponomarenko et al., ElliPro: A new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics. 9, 1–8 (2008).

Acknowledgments: We are grateful to Professors David Gray and members of the Gilbert Lab and MRC Human Genetics Unit for useful discussions throughout the project. We are grateful to Scott Neilson of the Edinburgh Genome Foundry for assistance with robotic cloning pipelines. We are grateful to the staff of the Scottish National Blood Transfusion Service Microbiology Reference Unit staff for provision of samples from Scottish blood donors. This research was conducted with material produced with the assistance of the Edinburgh Genome Foundry, a synthetic biology research facility specialising in the assembly of large DNA fragments at the University of Edinburgh.

Members of the Manchester Coronavirus Immune Response and Clinical Outcomes (CIRCO) consortium: Rohan Ahmed, Miriam Avery, Katharine Birchall, Evelyn Charsley, Alistair Chenery, Christine Chew, Richard Clark, Emma Connolly, Karen Connolly, Simon Dawson, Laura Durrans, Hannah Durrington, Jasmine Egan, Kara Filbey, Claire Fox, Helen Francis, Miriam Franklin, Susannah Glasgow, Nicola Godfrey, Kathryn J. Gray, Seamus Grundy, Jacinta Guerin, Pamela Hackney, Chantelle Hayes, Emma Hardy, Jade Harris, Anu John, Bethany Jolly, Verena Kästele, Gina Kerry, Sylvia Lui, Lijing Lin, Alex G. Mathioudakis, Joanne Mitchell, Clare Moizer, Katrina Moore, Stuart Moss, Syed Murtuza Baker, Rob Oliver, Grace Padden, Christina Parkinson, Michael Phuycharoen, Ananya Saha, Barbora Salcman, Nicholas A. Scott, Seema Sharma, Jane Shaw, Joanne Shaw, Elizabeth Shepley, Lara Smith, Simon Stephan, Ruth Stephens, Gael Tavernier, Rhys Tudge, Louis Wareing, Roanna Warren, Thomas Williams, Lisa Willmore, and Mehwish Younas.

Funding:

PKAK. is supported by a ECAT-Wellcome fellowship (223058/Z/21/Z).

NG is supported by MRC (MC_UU_00007/13).

JAM is a Lister Institute Research Fellow.

Lifearc, UKRI, Medical Research Scotland provided funding for development of the method.

The long COVID study was funded by the chief Scientist Office Scotland.

The CIRCO (Manchester) study was funded by the Wellcome Trust (202865/Z/16/Z).

The Oxford cohort was funded via the UK Vaccine Taskforce via NIHR grant to support the running of the Oxford ChAdOx1 nCoV-19 vaccine trial paid to the University of Oxford.

Author contributions:

Conceptualization: PKAK, DJK, JM, NG

Methodology: PKAK, DJK, DC, RK, KL, JM, NG

Computational/bioinformatic analyses: PKAK, MB, OF, LG, JM

Laboratory experiments: PKAK, KL, OF, SB, RK, DJK, NG

Edinburgh cohort: HW, SJ, KT

Oxford cohort: AM, JK, CD, TL, AP

Manchester cohort: MM, TH

Long Covid cohort: AC, LMcW

ELISA and statistical analyses: PKAK

Visualization: PKAK, MB

Funding acquisition: PKAK, DJK, JM, NG

Project administration: PKAK, OF, CD, TL, MM, TH, SJ, KT, AC, NG

Supervision: TL, JK, AP, TH, KT, SR, DC, JM, NG

Writing – original draft: PKAK

Writing – review & editing: PKAK, MB, AP, JK, AC, DJK, JAM, NG

All authors reviewed and approved the submitted manuscript.

Competing interests:

PKAK, DJK, JAM, and NG are inventors on a preliminary application for a patent filed 9th December 2021 by University of Edinburgh titled "Thermodynamic prediction, synthesis and prioritisation of immunogenic peptides" and have no other competing interests. AC is a paid editor of JNNP, unpaid president of functional neurological disorders society and gives independent testimony in court on a range of neuropsychiatric subjects on a roughly 50% claimant 50% defender split. LMcW is unpaid as a secretary of British Neuropsychiatry Association and gives independent testimony on court on a range of neuropsychiatric subjects. TL is named as an inventor on a patent application for a vaccine against SARS CoV-2 for an unrelated project. TL was a consultant to Vaccitech for an unrelated project. AJP is chair of the UK Department of Health and Social Care's Joint Committee on Vaccination and Immunisation but does not participate in the JCVI COVID19 committee. He was previously a member of WHO's SAGE. The University of Oxford has a partnership with AstraZeneca for the development of COVID19 vaccines. Other authors declare that they have no competing interests.

Data and materials availability: All data necessary are available in the text or the supplementary materials. Code to reproduce the analyses in the paper can be accessed from the first author's GitHub project repository (github.com/PKKearns/SARS2_Antibody_Profiling_Paper). Materials generated during the

course of this study: aliquots of purified GST-peptide fusion proteins for the prioritized peptides and bacterial/mammalian expression vectors are available on request.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S13

Tables S1 to S7

References (*93* - *106*)

Data Files S1 to S2