ORIGINAL ARTICLE

immunology

Alternative pathway dysregulation in tissues drives sustained complement activation and predicts outcome across the disease course in COVID-19

Matthew K. Siggins[1] | Kate Davies[2] | Rosie Fellows[3] | Ryan S. Thwaites[1] | J. Kenneth Baillie[4] | Malcolm G. Semple[5] | Peter J. M. Openshaw[1] | Wioleta M. Zelek[2] | Claire L. Harris[3] | B. Paul Morgan[2] | on behalf of the ISARIC4C Investigators

[1]National Heart & Lung Institute, Imperial College London, London, UK

[2]Division of Infection and Immunity and UK Dementia Research Institute Cardiff, Cardiff University, Cardiff, UK

[3]Translational & Clinical Research Institute, Newcastle University, Newcastle, UK

[4]Roslin Institute, University of Edinburgh, Edinburgh, UK

[5]NIHR Health Protection Research Unit, Institute of Infection, Veterinary, and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK

Correspondence
B. Paul Morgan, Henry Wellcome Building, Cardiff University Heath Park Campus, Cardiff CF14 4XN, UK.
Email: morganbp@cardiff.ac.uk

Abstract

Complement, a critical defence against pathogens, has been implicated as a driver of pathology in COVID-19. Complement activation products are detected in plasma and tissues and complement blockade is considered for therapy. To delineate roles of complement in immunopathogenesis, we undertook the largest comprehensive study of complement in COVID-19 to date, comprehensive profiling of 16 complement biomarkers, including key components, regulators and activation products, in 966 plasma samples from 682 hospitalized COVID-19 patients collected across the hospitalization period as part of the UK ISARIC4C (International Acute Respiratory and Emerging Infection Consortium) study. Unsupervised clustering of complement biomarkers mapped to disease severity and supervised machine learning identified marker sets in early samples that predicted peak severity. Compared to healthy controls, complement proteins and activation products (Ba, iC3b, terminal complement complex) were significantly altered in COVID-19 admission samples in all severity groups. Elevated alternative pathway activation markers (Ba and iC3b) and decreased alternative pathway regulator (properdin) in admission samples were associated with more severe disease and risk of death. Levels of most complement biomarkers were reduced in severe disease, consistent with consumption and tissue deposition. Latent class mixed modelling and cumulative incidence analysis identified the trajectory of increase of Ba to be a strong predictor of peak COVID-19 disease severity and death. The data demonstrate that early-onset, uncontrolled activation of complement, driven by sustained and progressive amplification through the alternative pathway amplification loop is a ubiquitous feature of COVID-19, further exacerbated in severe disease. These findings provide novel insights into COVID-19 immunopathogenesis and inform strategies for therapeutic intervention.

Matthew K. Siggins, Kate Davies and Rosie Fellows contributed equally to the work.

INTRODUCTION

Complement is a complex innate immune surveillance system, playing a key role in defence against pathogens. From early in the COVID-19 pandemic, the complement system has been implicated in the pathology of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection [1–3]. Accumulating evidence suggests that over-activation of the tightly controlled complement system triggers immune and inflammatory pathways that drive inflammation and tissue damage, leading to a vicious cycle of further complement activation and damage [4–7]. All three activation pathways, classical, lectin and alternative, have been implicated as triggers, and terminal pathway activation demonstrated in numerous studies. Complement over-activation likely contributes to COVID-19 pathology not only driving hyperinflammatory host responses but also exacerbating endothelial injury and hypercoagulability [7–9]. Complement activation products, notably C5a and the membrane attack complex (MAC), trigger and perpetuate inflammation by activating pro-inflammatory pathways in leukocytes, endothelial cells and other targets, while MAC directly causes endothelial cell damage and may also induce and activate cytotoxic T cells to exacerbate injury [10, 11]. Crosstalk with other pro-inflammatory cascades, for example, the kinin system, amplifies inflammation, while crosstalk with the coagulation system generates a pro-coagulant state that, in combination with endothelial injury, facilitates clotting [4–6, 12, 13].

Elevated levels of complement activation products and consumption of complement proteins in plasma, and abundant deposition of complement activation products in lung, kidney and other involved tissues, are reported in COVID-19 patients, particularly in severe disease [3, 14–20]. Plasma complement biomarker studies in COVID-19 published to date have been limited in terms of sample number, coverage of disease course and/or selection of analytes, providing sparse information on complement dysregulation in mild or early COVID [14, 15, 21–25]. The few studies that have tested in sequential samples from patients comprised small sample sets and measured individual analytes in fixed comparisons (e.g., hospitalized versus non-hospitalized, Intensive Care [ICU] versus non-ICU, or fatal versus non-fatal) [14–18]; none explored the diagnostic and/or prognostic value of measuring complement biomarkers across the disease course. Hence, there remains considerable uncertainty regarding how SARS-CoV-2 infection triggers complement over-activation, when in the disease course this occurs, which complement pathways drive over-activation and precisely how complement contributes to pathology across the disease course. This knowledge gap has important implications not only for diagnosis and prognosis but also for therapy; effective interventions targeting complement require an understanding of how, when and where complement activation occurs in the disease process and which activation pathways predominate. Indeed, although early studies testing available complement-blocking drugs in severe COVID-19 reported some remarkable outcomes [14, 21, 22, 26, 27]; none of the large phase II/III trials that followed met endpoint criteria [28, 29]. These failures were likely a consequence of the knowledge gap noted above; trial protocols included neither stratification of patients for therapy based on evidence of complement dysregulation nor evidence-based selection of complement targets.

These studies implicate complement dysregulation as a driver of pathology in COVID-19, but mechanistic details remain unclear. In particular, there is a lack of understanding of how complement dysregulation develops over the course of the infection. To clarify this critical element of pathogenesis, we conducted the largest comprehensive profiling of complement in infectious disease to date, assessing almost 1000 samples from hospitalized COVID-19 patients collected as part of the ISARIC4C study (https://isaric4c.net/) [30, 31]. Collectively, the data demonstrate that uncontrolled activation of complement, propagated by sustained and progressive alternative pathway amplification, occurs in COVID-19; they suggest that tissue deposition of complement complexes is an important component of severe disease. These findings significantly contribute to understanding how complement contributes to COVID-19 pathogenesis and have relevance for understanding the roles of complement in other infectious diseases.

RESULTS

Complement over-activation is apparent on admission in all severity groups, associated with severity at admission and predicts peak disease

To determine the relationship between levels of plasma complement biomarkers and disease severity in COVID-19, we used well-validated immunoassays to measure

TABLE 1 Patient characteristics for the ISARIC4C donors used in the study

| Characteristic | Overall, N = 682[a] | Peak severity | | | | Death, N = 142[a] | p-value[b] |
		Ward, N = 91[a]	Oxygen alone, N = 232[a]	NIV/HFNC, N = 80[a]	IMV, N = 137[a]		
Age	61 (51, 72)	64 (49, 77)	60 (50, 76)	58 (51, 65)	55 (46, 62)	66 (59, 75)	<0.001
(Unknown)	5	1	1	0	2	1	
Sex							<0.001
Female	249 (37%)	40 (44%)	102 (44%)	32 (40%)	41 (30%)	34 (24%)	
Male	430 (63%)	51 (56%)	130 (56%)	48 (60%)	95 (70%)	106 (76%)	
(Unknown)	3	0	0	0	1	2	
Obese	304 (64%)	45 (73%)	101 (64%)	31 (58%)	60 (58%)	67 (67%)	0.3
(Unknown)	206	29	75	27	33	42	
Any comorbidity	254 (53%)	38 (61%)	86 (55%)	26 (49%)	46 (44%)	58 (58%)	0.2
(Unknown)	206	29	75	27	33	42	
Cardiac disease	93 (20%)	20 (33%)	28 (18%)	4 (8.3%)	15 (15%)	26 (27%)	0.005
(Unknown)	227	30	80	32	38	47	
Diabetes	109 (23%)	14 (23%)	28 (18%)	12 (23%)	21 (20%)	34 (34%)	0.046
(Unknown)	206	29	75	27	33	42	
Pulmonary disease	53 (12%)	12 (20%)	20 (13%)	5 (10%)	6 (6.1%)	10 (10%)	0.11
(Unknown)	226	30	81	31	38	46	
Asthma	74 (16%)	10 (16%)	28 (18%)	10 (21%)	17 (17%)	9 (9.6%)	0.4
(Unknown)	229	30	80	32	39	48	
Kidney disease	49 (11%)	7 (12%)	17 (11%)	6 (12%)	8 (8.1%)	11 (12%)	>0.9
(Unknown)	230	31	81	32	38	48	
Corticosteroids	83 (23%)	3 (7.3%)	22 (18%)	9 (19%)	24 (33%)	25 (31%)	0.004
(Unknown)	318	50	108	33	65	62	
ICU	302 (44%)	1 (1.1%)	16 (6.9%)	44 (55%)	137 (100%)	104 (73%)	<0.001
(Unknown)	3	3	0	0	0	0	
Status							<0.001
Died	142 (22%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	142 (100%)	
Discharged alive	455 (69%)	79 (96%)	219 (96%)	76 (97%)	81 (64%)	0 (0%)	
On-going care	59 (9.0%)	3 (3.7%)	8 (3.5%)	2 (2.6%)	46 (36%)	0 (0%)	
(Unknown)	26	9	5	2	10	0	

(Continues)

TABLE 1 (Continued)

Characteristic	Overall, $N = 682$[a]	Peak severity					p-value[b]
		Ward, $N = 91$[a]	Oxygen alone, $N = 232$[a]	NIV/HFNC, $N = 80$[a]	IMV, $N = 137$[a]	Death, $N = 142$[a]	
Symptom onset to admission (days)	7.0 (3.0, 10.0)	3.0 (0.0, 7.0)	7.0 (3.0, 10.0)	7.0 (5.8, 10.0)	7.0 (5.0, 10.0)	7.0 (2.8, 10.0)	<0.001
(Unknown)	29	9	6	0	8	6	
Symptom onset to outcome (days)	18 (12, 27)	12 (8, 17)	14 (11, 20)	18 (12, 21)	33 (22, 50)	20 (13, 28)	<0.001
(Unknown)	51	14	9	2	20	6	
Admission to outcome (days)	10 (6, 19)	7 (3, 15)	7 (4, 12)	8 (5, 15)	25 (14, 43)	12 (8, 20)	<0.001
(Unknown)	31	9	3	2	16	1	

[a]Median (IQR); n (%).
[b]Kruskal-Wallis rank sum test; Pearson's Chi-squared test.

16 complement proteins chosen to include key components, regulators and activation products, and to interrogate activation and terminal pathways. Of the 750 patients recruited, 30 had a known symptom onset of over 28 days before admission or over 14 days after admission, and 38 lacked adequate outcome data; both these groups were excluded. From the remaining 682 cases, 966 plasma samples, plus samples from 49 healthy controls, a total of 1015 samples were measured (Table 1). The median duration of symptoms prior to recruitment and sample collection was 7 days, median symptom onset to outcome was 18 days and median admission to outcome was 10 days (Table 1). Corticosteroid use was significantly higher in more severe disease groups. Sample numbers for each biomarker are shown in Table S1. Donors were segregated based on severity into five clinical groups based on their peak disease severity as described in Methods (1, Ward; 2, Oxygen alone; 3, NIV/HFNC; 4, IMV; 5, Death); numbers in each severity group are in Table 1. Samples were further grouped based on the day of sampling post-admission (1–3; 4–7; 8–14; and Convalescent).

Admission (day 1–3) samples were available from 414 patients; data from these and the 49 healthy control samples were clustered using a hierarchical-k means hybrid approach and annotated with peak severity, daily severity (both on the 5-stage WHO Ordinal scale described above), ICU admission (Yes/No), disease duration at the time of sampling (from symptom onset; days), time from admission to sampling (days); time from sample to outcome (days), and length of hospital stay (days). Covariables were age, sex (Male/Female), presence of comorbidities (other than obesity) (Yes/No) and obesity (Yes/No; subjectively assessed by a clinician) for each patient. All comorbidities are listed in Methods. Healthy control samples ($n = 49$) were included as comparators (coloured green in Figure 1). Complement biomarker profiles clustered into four distinct patterns, apparent as separate hierarchical clusters in the heat map (Figure 1a). Cluster 4, defined by low levels of the activation markers Ba, iC3b and TCC and normal levels of other complement biomarkers, contained almost all healthy donor samples with only 15 COVID-19 samples. Cluster 3, defined by the highest levels of activation markers iC3b and TCC, moderately elevated levels of Ba and high levels of all components and regulators except C4 and FD, was associated with low levels of severe disease (22%), low ICU admissions (27%) and a low proportion of deaths (12%). Cluster 1, defined by moderately elevated levels of the activation markers Ba, iC3b and TCC, moderately reduced regulators clusterin and properdin and (compared to Cluster 3) reduced levels of other components and regulators, was associated with
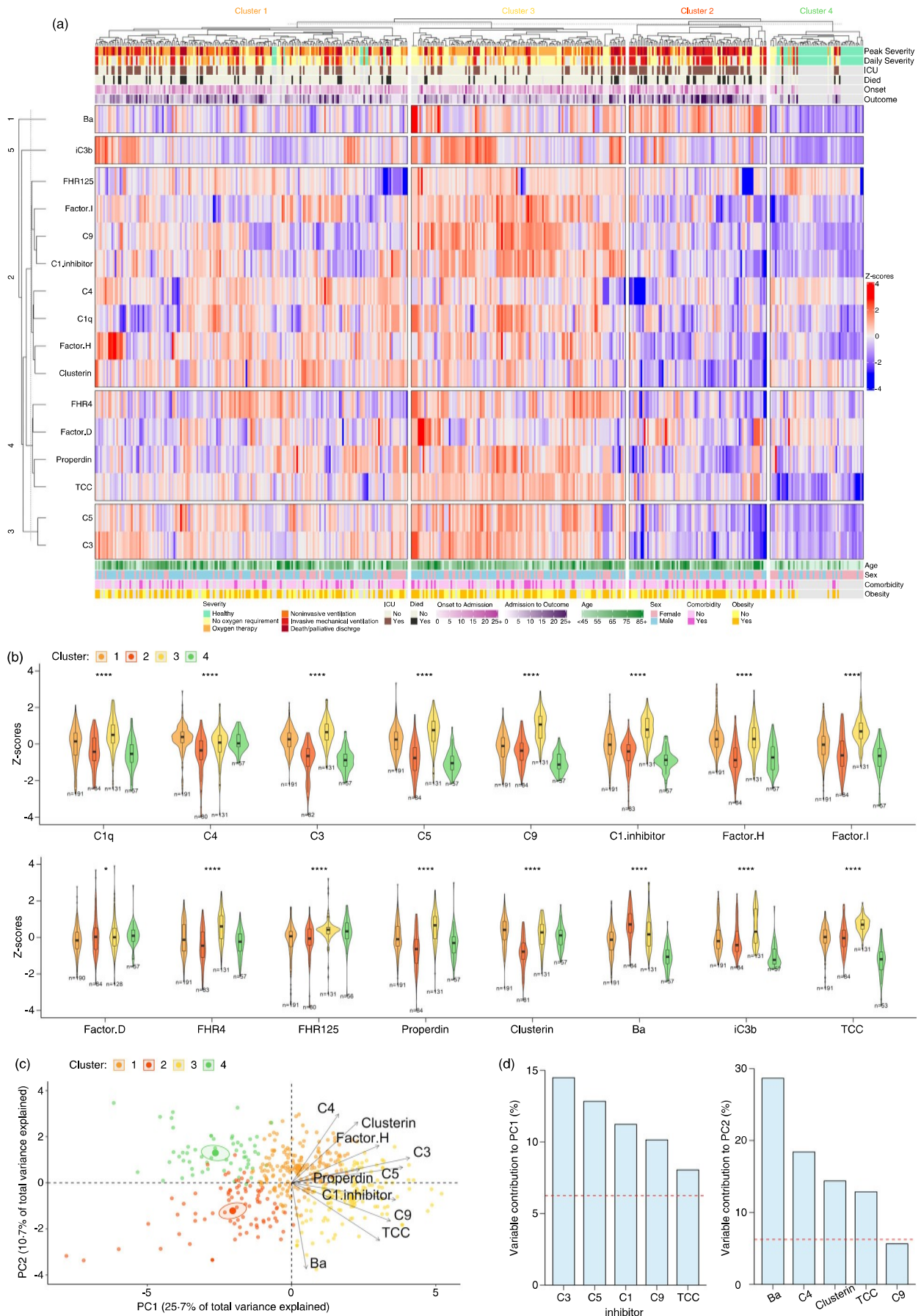
FIGURE 1 Legend on next page.

moderate levels of severe disease (33%), intermediate ICU admissions (36%) and deaths (17%). Cluster 2, defined by very high levels of the alternative pathway activation marker Ba but only moderately increased iC3b and TCC levels, very low levels of properdin and clusterin, and low levels of most other complement biomarkers including FH, FI, C1q, C4, C3 and C5, was associated with the most severe disease (63%), highest ICU admissions (62%) and deaths (28%). The Z scores for each biomarker in each disease severity group are shown in Figure 1b with clusters colour coded as in the figure key; significant differences between the clusters were evident for all biomarkers. Principal components analysis showed clear separation of healthy (Cluster 4; green, includes all healthy controls), mild-to-moderate disease (Clusters 3 and 1; yellow and orange respectively) and severe (Cluster 2; red) disease groups (Figure 1c). The contributions of the different biomarkers to each of the principal components in the analysis are shown (Figure 1d).

Clustering analysis revealed associations between complement analyte levels and disease severity; hence, we next used supervised machine learning to predict peak disease severity from admission samples and gain insight into the disease process by identifying key complement analytes with strong predictive power. Partial least squares discriminant analysis (PLS-DA) demonstrated clear separation of healthy from disease groups and separation, albeit with substantial overlap, between moderate and severe peak disease severity groups (Figure 2a, b). Healthy controls were characterized by low levels of C3, C5, C1inh and the activation products iC3b, Ba and TCC compared to the disease groups, with TCC providing the strongest differentiation of healthy from COVID-19 samples (Figure 2a, b, d). Moderate peak severity in COVID-19 was most strongly associated with elevated TCC and C3, and reduced clusterin and properdin, while severe peak disease was most strongly

associated with elevated Ba levels and reduced properdin and clusterin (Figure 2a, b). Cross-validated ROC analysis confirmed the clear separation of healthy from disease (AUC 0.989) and showed strong predictive power for separation of individual disease groups (AUC 0.826–0.841; Figure 2c). A PLS-DA relevance network showed that the three activation markers Ba, TCC and iC3b were negatively correlated with the healthy group, while Ba showed a positive correlation with severe peak disease. Strong negative correlation with properdin stood out as a predictor of severe peak disease, with negative correlations of TCC and Properdin defining Moderate peak disease (Figure 2e). As a complementary approach, cross-validated Random Forests analysis for all 16 complement variables was performed using admission (day 1–3) samples.

Overall, test-set accuracy in peak severity prediction was 84.6% and class-balanced predictive accuracy was 85.1% for severe disease (levels 6–8), 86.2% for moderate disease (levels 3–5) and 98.8% for healthy controls (Figure 2f). Subanalysis of variable importance demonstrated that TCC contributed most to the segregation of healthy from disease samples, while four complement variables (properdin>Ba>clusterin>C1q) provided most of the predictive power for peak disease severity in COVID-19 patients (Figure 2g). Partial dependence plots (PDP) were generated from the final random forests model to illustrate the relationship of combinations of the most informative variables to the prediction of severe peak disease while accounting for other variables (Figure 2h, i). Plotting Ba against TCC demonstrated that higher levels of both analytes increased the likelihood of severe peak disease with a synergistic relationship. Plotting either of these activation markers against properdin showed clear cut-off levels associated with increased likelihood of severe disease (Figure 2h). Scatter plots of raw data for all patients with values of plotted variables

FIGURE 1 Admission complement biomarker levels in patients hospitalized with COVID-19 show distinct response clusters that map to peak severity. (a) Unsupervised clustering heatmap of 16 complement biomarkers, including key activators, components and regulators in 414 plasma samples collected from patients hospitalized with COVID-19 within 3 days of admission, and 49 healthy control donors. Z scores were calculated from $\log_{10}$-transformed, scaled and centred values. A hybrid hierarchical k-means approach (hkmeans) was used to cluster patients by Euclidean distances using Ward's method. Four clusters are identified, labelled 1–4. Individual patient columns are annotated with key clinical and demographic information, indicated in the legend. Disease severity at the time of sampling (daily severity), peak severity, ICU admission, death as outcome, duration of illness before hospital admission, length of hospital stay, age, sex and comorbidity are shown. (b) Admission levels of 16 complement biomarkers from all patients separated based upon the four unsupervised clusters identified above (same colour codes). Violin plots show data distribution with overlayed boxplots indicating median and interquartile ranges. Numbers in each cluster are shown for all biomarkers. (c) Biplot of a principal component analysis (PCA) of Z scores of admission levels of 16 complement biomarkers from all patients included in the cluster analysis. Arrows show loadings of each listed complement biomarker. Points show PC1 and PC2 scores for individual patients, coloured as in the unsupervised clusters. Large points represent the centroids of each respective cluster with ellipses indicating the 95% confidence levels. *$p < 0.05$; ****$p < 0.0001$; Kruskal Wallis followed by Dunn's multiple comparisons post-tests with false discovery rates controlled using the Benjamini-Hochberg procedure. (d) Percent contributions of variables to PC1 and PC2 scores in the PCA analysis.
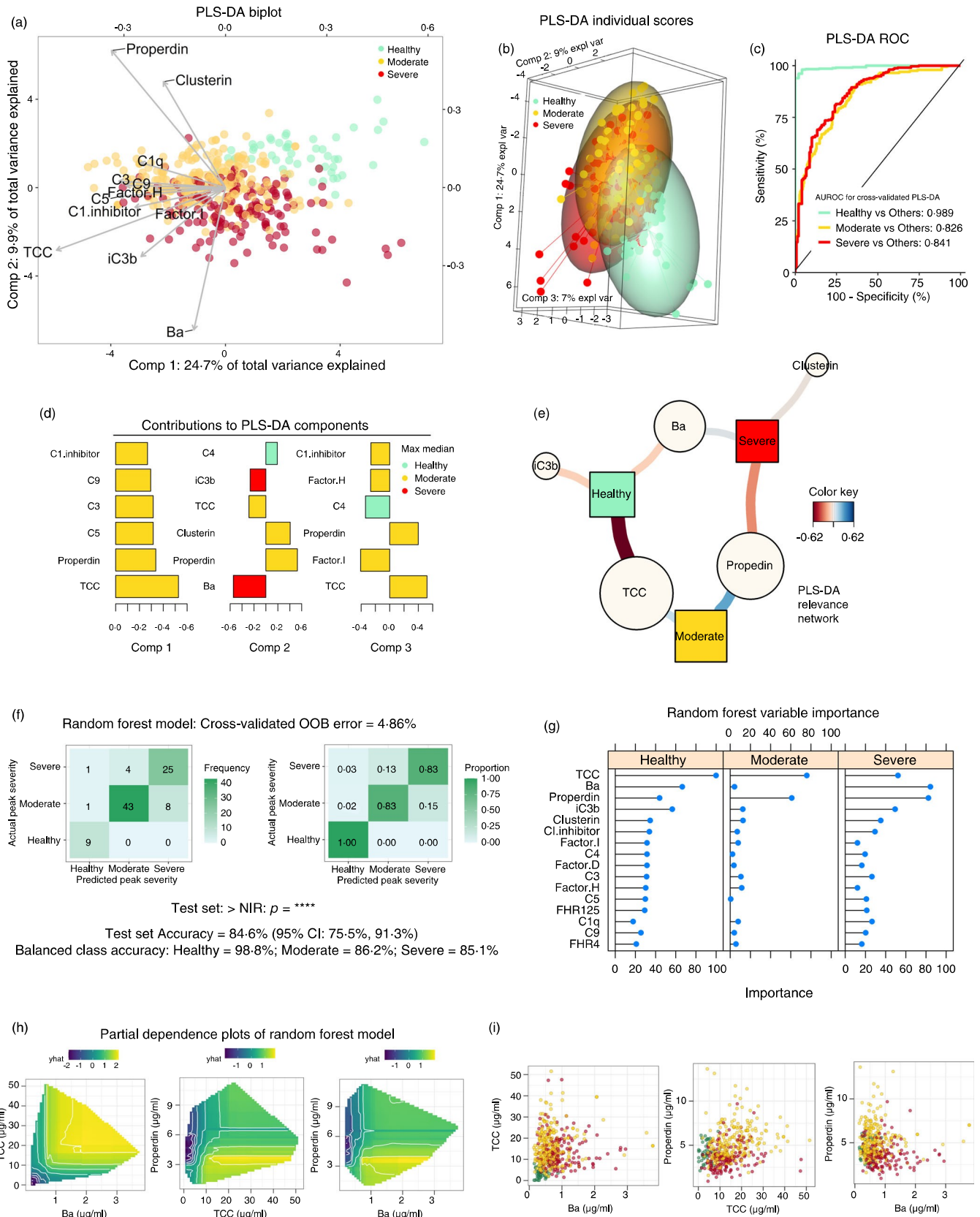
F I G U R E   2   Legend on next page.

coloured by actual peak severity are shown (Figure 2i); although these do not consider the impact of non-plotted complement analytes, they showed similar patterns. Absolute values for these key biomarkers and other patient characteristics in the different clusters are summarized in Tables S2 and S3.

Complement biomarkers distinguish disease severity groups across the disease course

Complement biomarkers measured at intervals post-admission were plotted in relation to peak severity (Figure 3) and severity at the time of sampling (Figure S1); for clarity of presentation, cases were divided into Moderate (score 3–5 on the WHO COVID-19 ordinal scale) and Severe (score 6–8 on the scale). Data were plotted in the different time interval groups (Figure 3a; Figure S1 a) and on a continuous time scale with the curve generated by Loess regression to demonstrate trends across hospitalization (Figure 3b; Figure S1 b).

All complement components tested, other than C4 and FD, were elevated in COVID-19 samples compared with healthy controls in both moderate and severe disease groups and at all timepoints, whether segregated based on peak severity (Figure 3a, b) or severity at the time of sampling (Figure S1 a, b). In admission (day 1–3) samples, levels of C1q, C3, C5 and C9 were all lower in severe disease compared to moderate, whether segregated on peak disease severity or severity at the time of sampling; on days 4–7 samples, these trends were less pronounced, significant only for C3; in day 7–14 samples, C1q, C3, C5 and C9 levels were all higher in the severe disease groups compared to moderate disease (Figure 3a; Figure S1 a).

Of the complement regulators C1inh, FH, FHR4, FHR125, FI and properdin, levels of C1inh, FH and clusterin were elevated in COVID-19 samples compared with healthy controls in all severity groups and at all timepoints, whether segregated based on peak severity (Figure 3a, b) or severity at the time of sampling (Figure S1 a, b); for other regulators, there was no consistent difference. In admission (day 1–3) samples, C1inh, FHR125, properdin and clusterin were all decreased with greater disease severity at the time of sampling and peak severity, with particularly large decreases for properdin and clusterin (Figure 3a; Figure S1 a); FH was significantly decreased in severe disease relative to moderate in the peak severity analysis (Figure 3a). In samples taken up to day 7 post-admission, peak disease severity and severity at the time of sampling were strongly associated with decreased properdin and clusterin; increased C1inh and FI were associated with severe disease in day 8–15 samples (Figure 3a; Figure S1 a). Loess curves of complement regulator levels highlighted decreased properdin and clusterin in severe disease across the time course, particularly in early samples, and the decrease in FI in later moderate disease samples (Figure 3b; Figure S1 b).

All measured complement activation products, Ba, iC3b and TCC, were elevated in COVID-19 samples compared with healthy controls in all severity groups and at all timepoints. Ba was markedly increased with increased disease severity at the time of sampling and with peak disease severity at all timepoints while iC3b levels were increased in severe disease but only in samples taken late in the disease course (Figure 3a; Figure S1 a). In contrast, TCC levels were reduced in more severe diseases assessed either at the time of sampling or at peak disease across the time course. Loess curves of activation product levels clearly illustrated the progressive increase in Ba over the

FIGURE 2 Supervised machine learning identifies key severity-associated complement biomarkers and predicts peak COVID-19 severity. Complement biomarkers were measured in 414 plasma samples collected from hospitalized COVID-19 patients within 3 days of admission, and 49 healthy control donors. (a) Z scores were generated and plotted with peak disease severity to generate a PLS-DA model to predict disease severity. Arrows in the biplot show loadings of each listed complement biomarker. Points show component 1 and 2 scores for individual patients, coloured by peak disease severity (green, healthy; yellow, moderate; red, severe). (b) 3D projection of components 1, 2, and 3 scores for all samples coloured by actual peak disease severity, with shading indicating the 95% confidence levels of the population concentration. (c) ROC curves and cross-validated AUC values from the PLS-DA classification illustrate the predictive power for peak severity prediction. (d) Loadings for the top six contributing complement biomarkers for PLS-DA components 1, 2, and 3. Colouring of bars indicates peak disease severity group with the maximum median value for each biomarker. (e) Relevance network showing key complement biomarkers contributing to disease severity prediction in the PLS-DA model. Colour key refers to connecting lines to show relative impact and direction of change (brown-to-red indicates association with low levels of biomarker; light blue-to-dark blue indicates association with high levels of biomarker). (f) Confusion matrix showing test set performance for prediction of peak severity using a random forests model trained on class-balanced and cross-validated data in the sample set. The cross-validated out-of-box (OOB) error rate was 4.86%. (g) Relative contributions of the different biomarkers to peak severity class prediction in the random forest model. (h) Partial dependence plots (PDP) for final random forest model, showing the average effects and interactions of plotted variables on prediction of severe peak disease while accounting for other variables. Colour bars above each plot indicate severity range from healthy (dark blue) to severe peak disease (yellow). (i) Scatter plots showing values of plotted variables for all patients, coloured by actual peak severity.
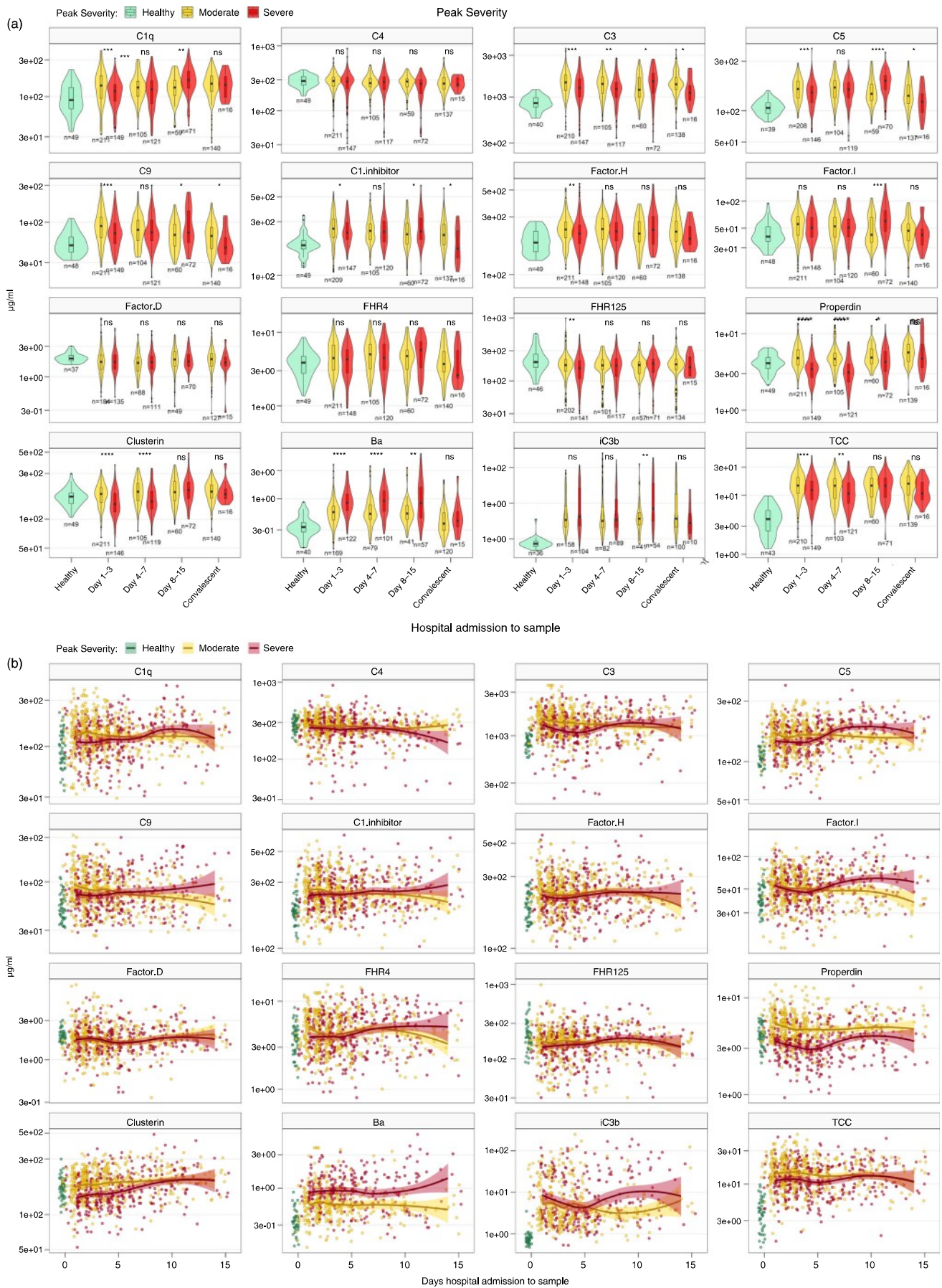
FIGURE 3 Legend on next page.

time course in severe disease and iC3b mirroring this rise in earlier samples while TCC levels were similar in the severity groups and constant across the time course (Figure 3b; Figure S1 b).

The analysis confirmed that Ba levels were robustly elevated in severe disease compared to moderate across the time course with an upward trajectory in severe disease. Properdin levels were consistently decreased across the disease course in severe disease compared to moderate while low clusterin levels were associated with severe disease in early disease but not later. These analyses also revealed an association of increased iC3b and FI levels with more severe peak disease in samples taken later post-admission.

Convalescent samples taken after hospital discharge were available from 168 patients (152 moderate peak disease, 16 severe peak disease); for most complement biomarkers, levels had normalized to near those in healthy controls and the peak disease severity groups were not significantly different (Figure 3a). Levels of C1q, C3, C9, C1inh, FH, iC3b and TCC remained high compared to healthy controls; C3, C5, C9 and C1inh were all higher in convalescent samples from the moderate peak disease group.

Markers of complement dysregulation correlate with outcome

At the time of analysis, outcome was known for 597 of the 682 COVID-19 patients included in the analysis; 455 had been discharged alive and 142 had died. The most informative complement biomarkers in those with death or discharge as an outcome were assessed for association with that outcome in samples taken across the time course (Figure 4a). Elevated Ba levels and decreased levels of clusterin and properdin in both admission samples and samples taken later post-admission were highly associated with death as outcome. Of the 302 patients treated in ICU, 104 died (Table 1); in a subanalysis of this ICU set, elevated Ba and decreased clusterin and

properdin levels were significantly associated with death as outcome in samples collected after day 4 post-admission (Figure S2a).

Loess curves demonstrated the trajectories of three biomarkers (Ba, clusterin, properdin) differed between outcome groups (Figure 4b; Figure S2b). For Ba in all samples, levels remained high or further increased in those with death as outcome and fell in those who survived; the widening gap between outcomes illustrated the power of this biomarker as a predictor of outcome (Figure 4b); in the ICU set widening of the gap between outcomes with time was obvious for Ba (increased in the death group), properdin and clusterin (both decreased in the death group) (Figure S2b). Longitudinal data of individual patients with repeated samples for the five complement biomarkers were also interrogated using spaghetti plots, colour-coded for outcome (survival, blue; death, red) and connected by lines to illustrate changes in biomarker levels across the time course for each individual (Figure 4c). For almost all individuals with death as outcome, Ba levels markedly increased over time, while for those who survived Ba levels were stable or fell over time.

Implication of the alternative pathway activation marker Ba as a predictor of outcome provoked us to develop a latent class linear mixed model to generate estimated trajectories for Ba across the disease time course (Figure 4d); two distinct classes of response were predicted in the final model, Class 1 (light blue) showed a downward Ba trajectory with time and was associated with moderate disease, while Class 2 (dark blue) showed a steeply upward Ba trajectory across the time course and was associated with severe disease and death. When plotted as individual patients and separated into peak disease severity groups, it is apparent that the mild and moderate peak disease groups contain mainly Class 1 Ba responses while Class 2 responses predominate in the more severe peak severity groups (Figure 4e). The Ba trajectory classes were also highly predictive of outcome as demonstrated by plotting cumulative incidence curves of death (red) and hospital discharge (blue) up to 28 days post

FIGURE 3 Complement biomarker levels across the admission correlate with peak disease severity and implicate alternative pathway over-activation as a driver of severe disease. (a) Plasma levels of 16 complement biomarkers measured from admission to 15 days post-admission in COVID-19 (735 samples) were grouped (days 1–3; 4–7; 8–15) and segregated by peak severity (moderate; severe). Numbers in each group are shown in the figure. Healthy control ($n = 49$) and convalescent (median 21 days post-discharge, IQR = 39 days; $n = 156$) samples were included for comparison. Violin plots show data distribution for moderate (yellow) and severe (red) peak disease severity with overlaid boxplots indicating median and interquartile range; significant differences are shown (*$p \leq 0.05$; **$p < 0.005$; ***$p < 0.001$; ****$p < 0.0001$; Kruskal Wallis test followed by Dunn's multiple comparisons post-test with false discovery rates controlled using the Benjamini-Hochberg procedure). (b) Plasma complement biomarker values in COVID-19 patients are shown on a continuous time scale with the curve generated by Loess regression demonstrating trends across the hospitalization and 95% confidence intervals (shaded areas) for moderate (yellow) and severe (red) peak disease. The healthy control values are shown in green at time zero for comparison.
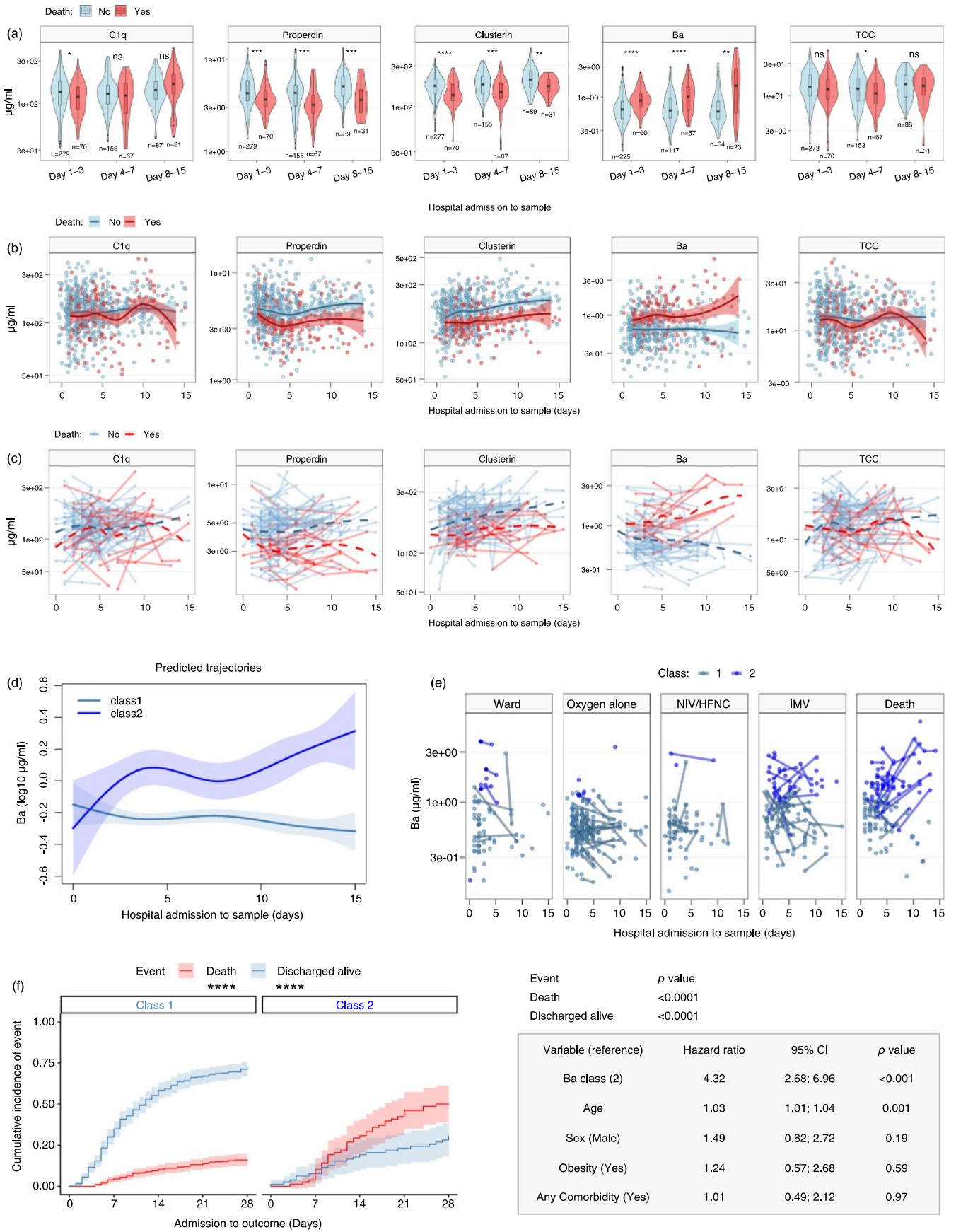
FIGURE 4 Legend on next page.

admission in COVID-19 patients (Figure 4f), with death and survival showing highly significant differences between classes (p value <0.0001 for both death and survival, Grey's modified Chi-squared test). Almost 75% of Class 1 (left plot) Ba trajectory cases had been discharged alive by day 28 post-admission compared to just 25% of Class 2 (right plot) Ba trajectory cases; over half of Class 2 Ba trajectory cases had died by day 28 compared to 15% of Class 1 cases. Competing risks regression, performed using a Fine-Grey proportional sub-distribution hazards model controlled for common patient characteristics (age, sex, obesity, and comorbidity), demonstrated that Ba trajectory class was a powerful predictor of death as outcome, with a large and highly statistically significant hazard ratio (HR 4.32, $p < 0.001$) (Figure 4f; inserted Table).

DISCUSSION

The COVID-19 pandemic has claimed over 6 million lives across the globe and continues to inflict severe costs. Better understanding of pathogenesis is critical for both the current pandemic and those to come. Evidence implicating complement in COVID-19 pathogenesis has accumulated over the course of the pandemic; however, a comprehensive understanding of how complement dysregulation is triggered, when in the disease process it occurs, which pathways of complement activation predominate and whether complement dysregulation is cause or consequence of the disease is lacking. Here, we demonstrate that complement dysregulation, and specific dysregulation of the alternative pathway, is a ubiquitous feature of severe COVID-19 across the disease course and predicts outcome.

The study has some limitations. We did not include assays specific for classical or lectin pathway activation; although C1q and C4 levels were measured and low levels suggest classical pathway engagement, inclusion of activation markers might have provided additional information on routes to dysregulation. More longitudinal sampling from individual patients may have improved the precision of reported complement trajectories; however, the large sample size with 485 of 966 (50.2%) analysed samples from patients with repeated samples, together with the application of established methods to combine repeated and single samples in the analysis, provided powerful coverage across the first 15 days of hospitalization with COVID-19. The study was limited to COVID-19 and did not include samples from patients with other forms of sepsis with ARDS; although it was not our intent here to look for COVID-19-specific markers, such samples might be informative.

Classical, lectin and alternative pathways have all been implicated as initiators and/or drivers of complement activation in COVID-19. Direct activation of the lectin pathway by SARS-COV-2 proteins and the presence of lectin pathway components in COVID-19 lung implies a role of lectin pathway in local pathology [19, 32]. SARS-CoV-2 antibody titres correlated with levels of the activation markers C3a and C5a in severe disease, suggesting an engagement of the classical pathway in

FIGURE 4 Progressive upward trajectories of alternative pathway activation markers are strongly and specifically associated with mortality in hospitalized COVID-19 patients. (a) Levels of the five most informative plasma complement biomarkers measured from admission to 15 days post-admission in COVID-19 were grouped (days 1–3; 4–7; 8–15) and segregated by final outcome (survival; death). Numbers in each group are shown in the figure. Violin plots show data distribution for survival (blue) and death (red) with overlaid boxplots indicating median and interquartile range. Significant differences are shown (*$p \leq 0.05$; **$p < 0.005$; ***$p < 0.001$; ****$p < 0.0001$; ns, $p > 0.5$. Kruskal Wallis followed by Dunn's multiple comparisons post-tests with false discovery rates controlled using the Benjamini-Hochberg procedure). (b) Data from A shown on a continuous time scale with a smooth curve generated by Loess regression demonstrating trends and 95% confidence intervals (shaded areas) for survival (blue) and death (red) as outcome. (c) Spaghetti plots showing longitudinal changes in levels of complement biomarkers in those individual patients with repeated measurements available for analysis. Samples from each patient are coloured based on outcome (survival, blue; death, red) and connected by lines; dashed lines indicate means for the survival (blue) and death (red) groups. (d) Latent class linear mixed model (LCLMM) showing estimated trajectories for the alternative pathway activation marker Ba and corresponding 95% confidence intervals during the first 15 days of hospitalization with COVID-19 for two distinct classes of response, Class 1 (light blue), associated with moderate disease, showing a downward trajectory, and Class 2 (dark blue), associated with severe disease and death, showing a steeply upward trajectory across the time course. Shaded areas represent 95% confidence intervals. (e) Plasma Ba levels in individual patient samples grouped by Ba trajectory class membership in the five different peak severity groups. Sequential samples from the same patient are connected by a line. The preponderance of Class 2 responses in the more severe peak severity groups is apparent. (f) Cumulative incidence curves of death (red) and hospital discharge (blue), grouped by Ba trajectory class membership (Class 1 on left, Class 2 on right), up to 28 days post admission in COVID-19 patients. Shaded areas represent 95% confidence intervals. Analysis was performed using Grey's modified Chi-squared test and Fine-Grey proportional sub-distribution hazards model. Obesity was clinician-defined and the comorbidities considered are described in detail in the methods. The appended table shows that a Ba Class 2 trajectory strongly predicts death as outcome.

later stages [33, 34]. The alternative pathway is distinct from the other complement activation pathways in that it is not a linear pathway; it is a positive feedback amplification loop that converts small activation events, whether initiated via the classical or lectin pathway or directly on activating surfaces, into large downstream effects [35]. This amplification property confers a high risk of uncontrolled activation, prevented in health by numerous fail-safe regulatory mechanisms. Failure of the tight regulation that holds the alternative pathway in check is a core feature of many complement dysregulation diseases [36]; indeed, the demonstration that SARS-CoV2 spike proteins directly activate the alternative pathway suggests that virus-triggered activation might be an initiator of dysregulation [37]. Several recent studies of complement biomarkers in COVID-19 have focussed attention on the alternative pathway [16, 17, 23–25]. Ma et al. showed an association of increased levels of FD, an essential alternative pathway enzyme, with severe disease and death as outcomes in hospitalized COVID-19 patients, provoking the suggestion that elevated FD levels were responsible for propagating alternative pathway activation [17]. Boussier et al. reported that levels of properdin, a critical stabilizer of the alternative pathway convertase, were reduced in severe COVID-19 and associated with the risk of requiring mechanical ventilation [23], although others reported the opposite – higher properdin levels in more severe disease and those with fatal outcome [24].

We measured levels of 16 complement components, regulators and activation products in 966 samples from 682 hospitalized COVID-19 patients, by far the largest study of complement dysregulation in COVID-19 to date. All samples were collected in the first wave of the pandemic before effective therapies were available. Cluster analysis showed that the complement biomarkers differentiated health from disease and segregated individuals into clusters with similar disease severity on admission and peak disease severity (Figure 1). Compared to healthy controls (all within Cluster 4 in the heat map), Z scores for ten of the measured complement analytes were elevated in COVID-19 admission samples; these included all three activation markers measured (Ba, iC3b, TCC), the regulators C1inh and FH, and all components except C4 and FD. The demonstration of increased levels of complement components in the face of activation and consumption reflects the fact that most are acute phase reactants, inflammation-driven increased synthesis compensating for increased consumption. Indeed, in Cluster 2, dominated by severe disease cases and poor outcome levels of components (C1q, C4, C3, C5, C9) and regulators (FH, FI, properdin, clusterin) were decreased compared to moderate disease (Clusters 1 and 3), suggesting that increased consumption outpaces production in

severe disease. The observation that C1q and C4 levels were reduced in the severe disease cluster implies increased classical pathway activation, perhaps driven by anti-SARS-CoV-2 antibodies [33, 34]. Z scores for the three activation markers were increased in all three disease clusters. Elevated levels of Ba, an alternative pathway-specific activation marker generated during the formation of the C3bBb convertase, were particularly evident in the most severe disease cluster, implying increased activity in the alternative pathway amplification loop. Notably, levels of the alternative pathway regulator FH were significantly lower in the severe disease cluster, suggesting that reduced FH exacerbates alternative pathway dysregulation in severe COVID-19.

PLS-DA and Random Forests analysis of data from samples taken early after admission confirmed the predictive value of the measured complement biomarkers in determining peak severity (Figure 2); the analyses showed separation of healthy from disease with a predictive value (healthy versus disease) of 0.989 (Figure 2c). This near-perfect segregation was driven largely by the three activation markers (Ba, iC3b, TCC); TCC, a marker of terminal pathway activation, was the best single discriminator between healthy and disease groups (Figure 2d, e). Optimal segregation of disease severity groups was provided by the three activation markers plus properdin and clusterin (Figure 2). Elevated Ba and decreased properdin and clusterin levels in admission samples were strongly predictive of severe peak disease. Elevated levels of TCC and/or C5a, indicative of terminal pathway activation, have been reported in several previous COVID-19 studies [14–18, 24, 25], provoking the suggestion that TCC, a stable and easy-to-measure marker, might aid diagnosis. However, we show that TCC, unlike other activation products measured, did not reflect disease severity, limiting its utility as a prognostic marker.

Three of the four complement analytes providing the best discrimination between severity groups in COVID-19 implicate the alternative pathway amplification loop, the activation products Ba and iC3b and the alternative pathway positive regulator properdin (Figure 2). In admission samples, Ba levels were significantly higher in those with severe compared to moderate disease, whether segregated based on severity at peak (Figure 3a) or at the time of sampling (Figure S1 a), whereas iC3b levels were not different between severity groups. Elevated plasma levels of activation fragments of FB (Ba, Bb) and/or C3 (C3a, C3b/iC3b) have been reported in severe COVID-19 patients in several small studies, further implicating alternative pathway dysregulation [16, 17, 23, 25]. Properdin levels in admission samples were significantly reduced in severe compared to moderate peak disease, demonstrating predictive value for disease outcome (Figure 3a), a

finding consistent with a recent small study (30 patients) showing that reduced properdin levels predicted the risk of requiring ventilation in severe COVID-19 [23]. Properdin levels remained low in severe COVID-19 across the disease course. Properdin stabilizes the alternative pathway convertase; reduced plasma levels would thus be anticipated to reduce alternative pathway activation. The likely explanation for this conundrum is that properdin acts mainly in tissues to localize and stabilize convertases [38, 39]; reduced plasma levels thus reflect properdin sequestration in tissues in severe disease where it contributes to alternative pathway over-activation. In samples collected later in the disease course, the observed increased Ba and decreased properdin levels persisted, demonstrating sustained alternative pathway over-activation in severe disease; iC3b was also increased in later severe peak disease samples (Figure 3a, b). Evidence of complement dysregulation, notably increased levels of the three activation markers, persisted even in convalescent samples; TCC levels remained high in moderate and severe peak disease convalescent samples (Figure 3a). This observation raises the possibility that complement activation may contribute to persistent post-COVID-19 symptoms.

Over-activation of the alternative pathway amplification loop also predicted the outcome. Increased Ba and decreased properdin were most predictive of death as outcome, whether measured on admission or at later time-points with increasingly divergent trajectories of change across the time course (Figure 4a, b, c). In patients admitted to ICU, admission values were not informative, expected as all ICU cases are, by definition, severe; however, in later samples, elevated Ba and decreased properdin levels were strong predictors of outcome (Figure S2a, b). Decreased levels of the alternative pathway regulator FH and increased levels of FD, an enzyme essential for alternative pathway convertase formation, were also associated with death as outcome in admission and/or early samples in the whole population and ICU subset, further implicating the pathway; the association of decreased FD with death as outcome replicates a recent study [17]. Implication of FH and FD as additional predictors of poor outcome illustrates the complexity underpinning alternative pathway dysregulation in severe COVID-19. Trajectories of Ba levels clearly separated severel from moderate disease across the time course (Figure 2d). Segregating individual patients based on peak disease severity or outcome clearly demonstrated the predictive value of the Ba trajectory model; class 2 trajectory was highly predictive of death as outcome (Figure 2e, f).

Of the other complement biomarkers measured, clusterin stood out. Decreased clusterin levels correlated with increased disease severity and were highly predictive of poor outcome in samples taken on admission and at all times post-admission. Clusterin is a multifunctional chaperone protein that binds and neutralizes potentially toxic moieties, including damage and dying cells; it inhibits the terminal pathway by binding precursor complexes and is incorporated in the TCC. It is therefore likely that clusterin is sequestered in tissue sites of complement dysregulation, both as a component of the TCC and directly binding damaged tissues [40]. Sequestration in tissues likely explains the association of reduced plasma clusterin levels with more severe disease in COVID-19; reduced plasma clusterin has been reported in sepsis and ascribed to sequestration [41]. Sequestration may also explain the observation that TCC levels were lower in more severe diseases in our study. TCC is likely trapped along with clusterin in the tissues; indeed, abundant deposition of TCC in kidneys and lungs has been reported in COVID-19 cases [3, 19, 20].

The data demonstrate that over-activation of the alternative pathway amplification loop, is a ubiquitous feature of COVID-19. The degree of activation correlates with severity and a rising trajectory of the alternative pathway-specific activation marker Ba predicts peak severity and outcome. The findings suggest that alternative pathway dysregulation is a driver of progression to severe disease and death in COVID-19. Complement activation is likely triggered early in the disease process in virus-infected lung tissue through a combination of lectin pathway activation on SARS-CoV-2 antigens, classical pathway activation driven by anti-viral antibodies, and direct activation on damaged tissue. The amplification loop of the alternative pathway is initiated via these activation pathways or directly by dead and damaged cells in the infected tissues. Properdin binding in tissues stabilizes the alternative pathway convertases and protects from regulators, leading to a vicious cycle of increased activation that drives severe disease. The rising trajectory of Ba in severe disease reflects this runaway alternative pathway activation and predicts outcome (see Graphic Abstract). Although complement dysregulation occurs in the infected tissues, complement activation products are released and will impact neutrophils and other immune cells, contributing to the systemic inflammatory response that typifies the disease [33, 34, 42, 43].

Our observations spotlight the alternative pathway amplification loop as the optimal target for therapy, demonstrate the need for early intervention and suggest a requirement for drugs acting locally in the lung to break the cycle of dysregulation and reduce local and systemic injury in COVID-19. Although all samples included in this study were collected in the first COVID-19 wave in 2020, these findings have relevance for later stages of the pandemic and new variants of SARS-CoV2;

understanding the drivers of pathology in the severe disease that typified the first wave prepares us for the future impact of waning immunity and escape variants. Indeed, the data are relevant beyond the current pandemic as similar mechanisms are likely to be involved in other severe viral respiratory diseases. A handful of small intervention studies targeting the amplification loop and trials of C3-blocking drugs are already in progress [44, 45]; however, targeting FB or FD would specifically block the alternative pathway-driven amplification essential for dysregulation while retaining classical and terminal pathway activities important in immune defence. Repurposing recently described orally active FD or FB inhibitors, already in trials for other conditions [46, 47], might prove effective; a trial of the small molecule FD inhibitor Danicopan in hospitalized COVID-19 patients, part of a large platform study (NCT04988035), has completed but is yet to report results. Patient selection and early intervention based on evidence of alternative pathway dysregulation would further enhance the design of trials and interventions, improving outcomes.

MATERIALS AND METHODS

Study design

This study is a component of the ISARIC WHO Clinical Characterization Protocol for Severe Emerging Infections in the UK (CCP-UK), an ongoing prospective cohort study of hospitalized patients with COVID-19, which is recruiting in 258 hospitals across England, Scotland, and Wales (National Institute for Health Research Clinical Research Network Central Portfolio Management System ID: 14152) (https://isaric4c.net/), a pre-approved pandemic preparedness study using a publicly available protocol with urgent public health research status. The ISARIC4C study collected comprehensive sets of data and samples from hospitalized patients with COVID-19 to facilitate better understanding of the disease process and to assist in developing effective treatments. It has enabled many outputs, including workstreams that better define clinical risk factors for disease severity and progression [31], demonstrate the contribution of host genetics to disease severity [48], and describe changes in immune and inflammatory markers [30]. The study continues to explore the impact of viral variants, environmental factors and host immune responses on disease severity. The protocol, including revisions, case report forms, patient information leaflets, consent forms and details of the Independent Data and Material Access Committee are available online (https://isaric4c.net/).

Study registration and approvals

The ISARIC WHO CCP-UK study was registered at https://www.isrctn.com/ISRCTN66726260 and designated an Urgent Public Health Research Study by the National Institute for Health Research UK. Ethical approval was given by the South Central - Oxford C Research Ethics Committee in England (Ref 13/SC/0149), the Scotland A Research Ethics Committee (Ref 20/SS/0028), and the WHO Ethics Review Committee (RPC571 and RPC572, 25 April 2013). Healthy controls were recruited specifically for the purpose of this study from healthy donors following informed consent at Newcastle University (REC reference 12/NE/0121).

Participants and clinical data collection

EDTA plasma samples were obtained from 750 individual patients hospitalized with PCR-proven SARS-CoV-2 infection during the first COVID-19 wave in the UK in 2020. Healthy control EDTA plasma samples were collected from 49 consented donors with no history of COVID-19 (mean age 37.1 years, 27 female). Blood was collected into EDTA tubes, separated promptly and plasma stored in aliquots at -80°C; freeze–thaw cycles were kept to a minimum and recorded for all sample sets. Thirty patients with a known symptom onset of over 28 days before admission or over 14 days after admission and 38 patients with unknown peak disease severity or outcome were excluded, leaving a total of 682 patients, 966 patient plasma samples and 49 control plasma samples for inclusion in the analyses. Samples were taken at multiple timepoints throughout the hospital stay; of the 682 patients available after exclusions, 481 had one sample available, 141 had two samples available, 46 had three samples available, and 14 had four or more samples available. These samples provided good coverage across the first 15 days of hospitalization. Distributions of age, sex, and ethnicity are shown in Table 1. Patients were stratified into five clinical groups based on their peak illness severity according to the World Health Organization COVID-19 ordinal scale [49]: (1) Ward; no oxygen requirement (Severity 3, 91 individuals); (2) Oxygen alone; patients requiring oxygen by face mask or nasal prongs (Severity 4, 232 individuals); (3) NIV/HFNC; patients requiring high-flow nasal cannula-delivered oxygen or non-invasive ventilation (Severity 5, 80 individuals); (4) IMV; patients requiring invasive mechanical ventilation (Severity 6/7, 137 individuals); and (5) Death; fatal outcome due to COVID-19 (Severity 8, 142 individuals) (Table 1). For clarity in some analyses, clinical groups 1–3 were combined as Moderate and groups 4 and 5 as

Severe. For analyses across the time course, samples were grouped as 1–3, 4–7 and 8–14 days based on the time of sampling post-admission following a protocol harmonized with international investigators to allow meaningful comparison of results between studies [50]. Severity at the time of sampling was also captured using the same WHO ordinal scale definitions. Convalescent samples (post-discharge) were obtained from 168 patients. ICU admission was recorded for 302 patients (43.8%), the large majority (95%) from the severe disease categories; outcomes were captured separately in the ICU groups (Table 1).

A prespecified case report form was used to collect data on patient characteristics, treatments received in the hospital and outcomes; the core data are captured in Table 1. A modified Charlson comorbidity index was used to define comorbidities, and obesity was clinician-defined. 'Any comorbidity' included diabetes mellitus, chronic cardiac disease, chronic pulmonary disease (excluding asthma), asthma (physician-diagnosed), chronic kidney disease, chronic hematologic disease, malignant neoplasm, dementia, or moderate–severe liver disease. The median and range of 1. duration of symptoms prior to admission and sample collection; 2. duration of symptoms from onset to outcome (discharge or death); 3. duration of symptoms from admission to outcome, were recorded for the whole group and for the different disease severity groups (Table 1).

Immunoassays

Sixteen complement biomarkers including components, regulators and activation products, were measured. Four analytes (C3, C5, iC3b, Ba) were quantified using in-house assays on the MSD platform (Mesoscale Diagnostics, Rockville, Maryland, USA); standard or streptavidin plates were used and analysed on an SQ120 Quickplex instrument. MSD provided the dynamic range necessary to capture accurate levels in healthy controls and patients at the same sample dilutions. Assays for C3 and C5 were designed to specifically measure only intact proteins and not the activation products (C3a/C3b/iC3b or C5a/C5b respectively), an important distinction from routine assays for C3 and C5. The remaining twelve biomarkers (C1q, C4, C9, C1 inhibitor (C1inh), factor H (FH), factor H-related protein 4 (FHR4), factor H-related proteins 1, 2 and 5 (FHR125), factor I (FI), properdin, factor D (FD), clusterin, terminal complement complex (TCC)) were measured using validated in-house ELISA. The FHR125 assay measures the total concentrations of the three homologous FHRs 1, 2 and 5; signal is dominated by the abundant FHR1. Three activation markers were

measured, iC3b, Ba and TCC. C3a and C5a were not measured because the supplied samples had been subject to either one or two freeze–thaw cycles during collection and distribution, shown in preliminary stability studies to strongly affect measured levels of C3a and C5a in an unpredictable manner; other biomarkers measured, including iC3b, Ba and TCC, were not significantly impacted. Antibodies used were either from commercial sources or made in-house; standards were the relevant pure proteins, either in-house or commercial (Table S1). Plasma dilutions for each biomarker were established in preliminary experiments. All assays passed stringent quality control tests, including measurement of intra- and inter-assay coefficients of variation [51]. Samples failing quality control for a specific biomarker were excluded from the analysis for that biomarker. Prior to testing, samples were mixed 1:1 with 1% triton x-100 in PBS to eliminate the risk of live virus [52].

Statistical analyses

Statistical analyses used R version 4.1 with R studio version 1.4.1717 and additional packages used, not part of base R, are listed alongside methodical details below. 'doParallel' and 'parallel' were used for parallel computing of some more intensive statistical calculations. All visualization used package 'ggplot2' unless otherwise stated and 'tidyverse' packages and 'ggpubr' were used to enhance figure presentation. Distribution of complement analyte data were assessed by quantile-quantile ('qqplotr') and density plots, as well as D'Agostino-Pearson ('fBasics') and Shapiro–Wilk's normality tests. Non-parametric two-way analyses were performed using Wilcoxon rank sum tests ('rstatix'). Comparisons of more than two groups were performed with Kruskal-Wallis tests followed by Dunn's test for multiple comparisons of patient groups ('rstatix'). False discovery rates were controlled using the Benjamini-Hochberg protocol.

An unsupervised clustered heatmap of complement analyte data used log_{10}-transformed, scaled and centred values were generated using 'ComplexHeatmap' and 'factoextra' from samples within the first 3 days post-admission. Clustering tendency was assessed by the Hopkins statistic and inspection of a dissimilarity matrix ('hopkins', 'factoextra'). Euclidean distances were clustered using Ward's minimum variance method and patients were clustered using a hybrid hierarchical k-means approach; complement analytes were clustered using a model-based clustering method ('mclust'). For model-based clustering, bayesian information criterion was used for model selection. Agglomerative and divisive hierarchical clustering, density-based and hierarchical

density-based hybrid clustering ('fpc' and 'dbscan'), K-means, spherical K-means ('skmeans'), and K-medoids ('cluster' and 'fpc') clustering approaches were also assessed. The final approaches and models were chosen based on the biological relevance, quality, and consistency of clusters, as well as consideration of cluster separation by principal component scores and values of key complement analytes. Stability and internal validation metrics and clustering statistics were assessed using 'clvalid', 'NbClust', 'fpc', and 'factoextra'.

Principal components analysis (PCA) and partial least squares discriminant analysis (PLS-DA) both used $\log_{10}$-transformed, scaled and centred complement analyte data within the first 3 days post-admission. PCA was performed using 'FactoMineR' and was visualized using factoextra. PLS-DA was performed using 'mixomics' and visualized using 'mixomics' and 'rgl'. 10-fold cross-validation with 50 repeats was used to demonstrate that a model using 8 PLS-DA components and Mahalanobis distance gave optimal performance, based on consideration of error rate, class-specific error rate, class-balanced error rate and complementary classification performance results from a receiver operating characteristic area under the curve.

For Random Forests, complement analyte data from within the first 3 days post-admission were randomly split 80:20 into training and testing datasets. Healthy, Moderate, and severe peak severity classes were balanced in the training dataset using 'ROSE'. Variable selection using 'Boruta' shadow features demonstrated all 16 complement analytes had predictive value in classifying peak severity. The 16-analyte model was tuned and trained on the training test set with 5 iterations of 10-fold cross-validation using 'caret' and 'randomForest', with final parameters of 2500 trees and a 'mtry' of 4 (number of randomly selected variables at each split) giving optimal performance. Variable importance for prediction of each peak disease severity class was calculated using 'caret'. Partial dependence plots (PDP) of the final cross-validated random forest model were generated using 'pdp' to provide intuitive visualization of potentially complex variable relationships on a target variable (in this case peak disease severity). PDP display the average effects of the two plotted complement analytes on the target variable (severe peak disease severity), for every value of the two plotted complement analytes in the training dataset, while accounting for the average effects of all other complement analytes. 'yhat' represents the relative contribution of the two plotted complement analytes on the prediction of severe peak disease severity, with positive values meaning prediction of severe peak disease severity is more likely for the corresponding values of plotted complement analytes. Negative yhat values mean prediction of severe peak disease severity is less likely, and a yhat of zero implies there is no average impact on the prediction of severe peak disease severity according to the model [53].

For multivariate analyses, missing values of complement (totalling less than 5% of the data set) were imputed by predictive mean matching using the Multivariate Imputation by Chained Equations (MICE) package (https://www.rdocumentation.org/packages/mice). These imputed data were not used for univariate presentations or analyses.

Latent class linear mixed models (LCLMM) were created using 'lcmm'. Trajectories of Ba over time from admission with varying numbers of latent classes were generated using linear or various spline-fitted ('splines') models with random intercepts and random slopes at the patient level using data from samples up to 15 days post-admission. Selection of the best model (2 latent classes with a natural spline function on admission to sample) was based on consideration of posterior classification scores, class sizes, and Bayesian information criteria.

Cumulative incidence curves of the competing events of death and hospital discharge of COVID-19 patients up to 28 days post-admission (grouped by Ba trajectory class membership from the LCLMM model) were generated with 'cmprsk' and visualized with 'survminer'. Grey's modified Chi-squared test ('cmprsk') was used to compare survival and death curves between Ba trajectory classes. Competing risks regression was performed using a Fine-Grey proportional subdistribution hazards model ('cmprsk') to assess the impact of Ba trajectory class and other covariates (age, sex, comorbidity, obesity, and onset to admission) on outcomes. The final model (Ba trajectory class, age, sex, comorbidity and obesity) was selected based on Bayesian information criteria. Schoenfeld residuals and covariate-time interactions were assessed to confirm the proportional hazard sub-distribution assumption.

AUTHOR CONTRIBUTIONS

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Matthew K. Siggins https://orcid.org/0000-0003-2504-6518
Kate Davies https://orcid.org/0000-0002-9807-1231
Peter J. M. Openshaw https://orcid.org/0000-0002-7220-2555
Wioleta M. Zelek https://orcid.org/0000-0002-2230-3550
B. Paul Morgan https://orcid.org/0000-0003-4075-7676

REFERENCES

1. Risitano AM, Mastellos DC, Huber-Lang M, Yancopoulou D, Garlanda C, Ciceri F, et al. Complement as a target in COVID-19? Nat Rev Immunol. 2020;20(6):343–4. https://doi.org/10.1038/s41577-020-0320-7
2. Java A, Apicelli AJ, Liszewski MK, Coler-Reilly A, Atkinson JP, Kim AH, et al. The complement system in COVID-19: friend and foe? JCI Insight. 2020;5(15):e140711. https://doi.org/10.1172/jci.insight.140711
3. Noris M, Benigni A, Remuzzi G. The case of complement activation in COVID-19 multiorgan impact. Kidney Int. 2020;98(2):314–22. https://doi.org/10.1016/j.kint.2020.05.013
4. Afzali B, Noris M, Lambrecht BN, Kemper C. The state of complement in COVID-19. Nat Rev Immunol. 2022;22:77–84. https://doi.org/10.1038/s41577-021-00665-1
5. Ghebrehiwet B, Peerschke EI. Complement and coagulation: key triggers of COVID-19–induced multiorgan pathology. J Clin Invest. 2020;130(11):5674–6. https://doi.org/10.1172/JCI142780
6. Fletcher-Sandersjöö A, Bellander BM. Is COVID-19 associated thrombosis caused by overactivation of the complement cascade? A literature review. Thromb Res. 2020;194:36–41. https://doi.org/10.1016/j.thromres.2020.06.027
7. Chauhan AJ, Wiffen LJ, Brown TP. COVID-19: a collision of complement, coagulation and inflammatory pathways. J Thromb Haemost. 2020;18(9):2110–7. https://doi.org/10.1111/jth.14981
8. Perico L, Benigni A, Casiraghi F, Ng LFP, Renia L, Remuzzi G. Immunity, endothelial injury and complement-induced coagulopathy in COVID-19. Nat Rev Nephrol. 2021;17(1):46–64. https://doi.org/10.1038/s41581-020-00357-4
9. Cugno M, Meroni PL, Gualtierotti R, Griffini S, Grovetti E, Torri A, et al. Complement activation and endothelial perturbation parallel COVID-19 severity and activity. J Autoimmun. 2021;116:102560. https://doi.org/10.1016/j.jaut.2020.102560
10. Hughes J, Nangaku M, Alpers CE, Shankland SJ, Couser WG, Johnson RJ. C5b-9 membrane attack complex mediates endothelial cell apoptosis in experimental glomerulonephritis. Am J Physiol Renal Physiol. 2000;278(5):F747–57. https://doi.org/10.1152/ajprenal.2000.278.5.F747
11. Xie CB, Jane-Wit D, Pober JS. Complement membrane attack complex: new roles, mechanisms of action, and therapeutic targets. Am J Pathol. 2020;190(6):1138–50. https://doi.org/10.1016/j.ajpath.2020.02.006
12. Lupu F, Keshari RS, Lambris JD, Coggeshall KM. Crosstalk between the coagulation and complement systems in sepsis. Thromb Res. 2014;133(Suppl 1(01)):S28–31. https://doi.org/10.1016/j.thromres.2014.03.014
13. Kenawy HI, Boral I, Bevington A. Complement-coagulation cross-talk: a potential mediator of the physiological activation of complement by low pH. Front Immunol. 2015;6(6):215. https://doi.org/10.3389/fimmu.2015.00215
14. Zelek WM, Cole J, Ponsford MJ, Harrison RA, Schroeder BE, Webb N, et al. Complement inhibition with the C5 blocker LFG316 in severe COVID-19. Am J Respir Crit Care Med. 2020;202(9):1304–8. https://doi.org/10.1164/rccm.202007-2778LE
15. Holter JC, Pischke SE, de Boer E, Lind A, Jenum S, Holten AR, et al. Systemic complement activation is associated with respiratory failure in COVID-19 hospitalized patients. Proc Natl Acad Sci U S A. 2020;117(40):25018–25. https://doi.org/10.1073/pnas.2010540117
16. de Nooijer AH, Grondman I, Janssen NAF, Netea MG, Willems L, van de Veerdonk FL, et al. RCI-COVID-19 study group. Complement activation in the disease course of coronavirus disease 2019 and its effects on clinical outcomes. J Infect Dis. 2021;223(2):214–24. https://doi.org/10.1093/infdis/jiaa646
17. Ma L, Sahu SK, Cano M, Kuppuswamy V, Bajwa J, McPhatter J, et al. Increased complement activation is a distinctive feature of severe SARS-CoV-2 infection. Sci Immunol. 2021;6(59):eabh2259. https://doi.org/10.1126/sciimmunol.abh2259
18. Henry BM, Szergyuk I, de Oliveira MHS, Lippi G, Benoit JL, Vikse J, et al. Complement levels at admission as a reflection

of coronavirus disease 2019 (COVID-19) severity state. J Med Virol. 2021;93(9):5515–22. https://doi.org/10.1002/jmv.27077

19. Pfister F, Vonbrunn E, Ries T, Jäck HM, Überla K, Lochnit G, et al. Complement activation in kidneys of patients with COVID-19. Front Immunol. 2021;11:594849. https://doi.org/10.3389/fimmu.2020.594849

20. Niederreiter J, Eck C, Ries T, Hartmann A, Märkl B, Büttner-Herold M, et al. Complement activation via the lectin and alternative pathway in patients with severe COVID-19. Front Immunol. 2022;13:835156. https://doi.org/10.3389/fimmu.2022.835156

21. Diurno F, Numis FG, Porta G, Cirillo F, Maddaluno S, Ragozzino A, et al. Eculizumab treatment in patients with COVID-19: preliminary results from real life ASL Napoli 2 Nord experience. Eur Rev Med Pharmacol Sci. 2020;24:4040–7.

22. Mastaglio S, Ruggeri A, Risitano AM, Angelillo P, Yancopoulou D, Mastellos DC, et al. The first case of COVID-19 treated with the complement C3 inhibitor AMY-101. Clin Immunol. 2020;215:108450.

23. Boussier J, Yatim N, Marchal A, Hadjadj J, Charbit B, El Sissy C, et al. Severe COVID-19 is associated with hyperactivation of the alternative complement pathway. J Allergy Clin Immunol. 2022;149(2):550–556.e2. https://doi.org/10.1016/j.jaci.2021.11.004

24. Alosaimi B, Mubarak A, Hamed ME, Almutairi AZ, Alrashed AA, AlJuryyan A, et al. Complement anaphylatoxins and inflammatory cytokines as prognostic markers for COVID-19 severity and in-hospital mortality. Front Immunol. 2021;12:668725. https://doi.org/10.3389/fimmu.2021.668725

25. Leatherdale A, Stukas S, Lei V, West HE, Campbell CJ, Hoiland RL, et al. Persistently elevated complement alternative pathway biomarkers in COVID-19 correlate with hypoxemia and predict in-hospital mortality. Med Microbiol Immunol. 2022;211:1–12. https://doi.org/10.1007/s00430-021-00725-2

26. Annane D, Heming N, Grimaldi-Bensouda L, Frémeaux-Bacchi V, Vigan M, Roux AL, et al. Eculizumab as an emergency treatment for adult patients with severe COVID-19 in the intensive care unit: a proof-of-concept study. EClinicalMedicine. 2020;28:100590. https://doi.org/10.1016/j.eclinm.2020.100590

27. Mastellos DC, da Silva BGP P, BAL F, Fonseca NP, Auxiliadora-Martins M, Mastaglio S, et al. Complement C3 vs C5 inhibition in severe COVID-19: early clinical findings reveal differential biological efficacy. Clin Immunol. 2020;220:108598. https://doi.org/10.1016/j.clim.2020.108598. Epub 2020 Sep 19. PMID: 32961333.

28. Smith K, Pace A, Ortiz S, Kazani S, Rottinghaus S. A phase 3 open-label, randomized, controlled study to evaluate the efficacy and safety of intravenously administered Ravulizumab compared with best supportive care in patients with COVID-19 severe pneumonia, acute lung injury, or acute respiratory distress syndrome: a structured summary of a study protocol for a randomised controlled trial. Trials. 2020;21(1):639. https://doi.org/10.1186/s13063-020-04548-z

29. Vlaar APJ, de Bruin S, Busch M, Timmermans SAMEG, van Zeggeren IE, Koning R, et al. Anti-C5a antibody IFX-1 (vilobelimab) treatment versus best supportive care for patients with severe COVID-19 (PANAMO): an exploratory, open-label,

phase 2 randomised controlled trial. Lancet Rheumatol. 2020;2(12):e764–73. https://doi.org/10.1016/S2665-9913(20)30341-6

30. Thwaites RS, Sanchez Sevilla Uruchurtu A, Siggins MK, Liew F, Russell CD, Moore SC, et al. Inflammatory profiles across the spectrum of disease reveal a distinct role for GM-CSF in severe COVID-19. Sci Immunol. 2021 Mar 10;6(57):eabg9873. https://doi.org/10.1126/sciimmunol.abg9873

31. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. BMJ. 2020 May;22(369):m1985. https://doi.org/10.1136/bmj.m1985

32. Ali YM, Ferrari M, Lynch NJ, Yaseen S, Dudler T, Gragerov S, et al. Lectin pathway mediates complement activation by SARS-CoV-2 proteins. Front Immunol. 2021;12:714511. https://doi.org/10.3389/fimmu.2021.714511

33. Kim DM, Kim Y, Seo JW, Lee J, Park U, Ha NY, et al. Enhanced eosinophil-mediated inflammation associated with antibody and complement-dependent pneumonic insults in critical COVID-19. Cell Rep. 2021;37(1):109798. https://doi.org/10.1016/j.celrep.2021.109798. Epub 2021 Sep 20. PMID: 34587481; PMCID: PMC8450316.

34. Jarlhelt I, Nielsen SK, Jahn CXH, Hansen CB, Pérez-Alós L, Rosbjerg A, et al. SARS-CoV-2 antibodies mediate complement and cellular driven inflammation. Front Immunol. 2021;12:767981. https://doi.org/10.3389/fimmu.2021.767981. PMID: 34804055; PMCID: PMC8596567.

35. Harrison RA. The properdin pathway: an 'alternative activation pathway' or a 'critical amplification loop' for C3 and C5 activation? Semin Immunopathol. 2018;40(1):15–35. https://doi.org/10.1007/s00281-017-0661-x

36. Zipfel PF, Heinen S, Józsi M, Skerka C. Complement and diseases: defective alternative pathway control results in kidney and eye diseases. Mol Immunol. 2006;43(1–2):97–106. https://doi.org/10.1016/j.molimm.2005.06.015

37. Yu J, Yuan X, Chen H, Chaturvedi S, Braunstein EM, Brodsky RA. Direct activation of the alternative complement pathway by SARS-CoV-2 spike proteins is blocked by factor D inhibition. Blood. 2020;136(18):2080–9. https://doi.org/10.1182/blood.2020008248

38. Lesher AM, Nilsson B, Song WC. Properdin in complement activation and tissue injury. Mol Immunol. 2013;56(3):191–8. https://doi.org/10.1016/j.molimm.2013.06.002

39. Chen JY, Cortes C, Ferreira VP. Properdin: a multifaceted molecule involved in inflammation and diseases. Mol Immunol. 2018;102:58–72. https://doi.org/10.1016/j.molimm.2018.05.018

40. Silkensen JR, Schwochau GB, Rosenberg ME. The role of clusterin in tissue injury. Biochem Cell Biol. 1994;72(11–12):483–8. https://doi.org/10.1139/o94-065

41. Garcia-Obregon S, Azkargorta M, Seijas I, Pilar-Orive J, Borrego F, Elortza F, et al. Identification of a panel of serum protein markers in early stage of sepsis and its validation in a cohort of patients. J Microbiol Immunol Infect. 2018;51(4):465–72. https://doi.org/10.1016/j.jmii.2016.12.002

42. Georg P, Astaburuaga-García R, Bonaguro L, Brumhard S, Michalick L, Lippert LJ, et al. Complement activation induces excessive T cell cytotoxicity in severe COVID-19. Cell. 2022;185(3):493–512.e25. https://doi.org/10.1016/j.cell.2021.12.040