

1 **Elevated plasma Complement Factor H Regulating Protein 5 is associated with venous**  
2 **thromboembolism and COVID-19 severity**

3  
4 **One sentence summary:** Elevated CFHR5 plasma concentration is associated with diagnosis  
5 and risk of venous thromboembolism, and with short-term prognosis in hospitalised COVID-19  
6 patients.

7  
8 Laura Sanchez-Rivera<sup>1</sup> †, Maria Jesus Iglesias<sup>1,2,3</sup> †, Manal Ibrahim-Kosta<sup>4</sup>, Julia Barbara Kral-  
9 Pointner<sup>5</sup>, Sebastian Havervall<sup>6</sup>, Louisa Goumidi<sup>4</sup>, Maria Farm<sup>5,7</sup>, Gaëlle Munsch<sup>8</sup>, Marine  
10 Germain<sup>8,9</sup>, Philip Smith<sup>10,11</sup>, Mun-Gwan Hong<sup>1</sup>, Pierre Suchon<sup>4</sup>, Clément Naudin<sup>1,3</sup>, Anne  
11 Boland<sup>12,13</sup>, David M Smadja<sup>14,15</sup>, Margareta Holmström<sup>16</sup>, Maria Magnusson<sup>5,16,17</sup>, Angela  
12 Silveira<sup>10</sup>, Mathias Uhlén<sup>1</sup>, Thomas Renné<sup>18,19,20</sup>, Angel Martinez-Perez<sup>21</sup>, Joseph Emmerich<sup>22</sup>,  
13 Jean-Francois Deleuze<sup>23,24,25</sup>, Jovan Antovic<sup>5,7</sup>, Alice Assinger<sup>26</sup>, Jose Manuel Soria  
14 Fernandez<sup>21</sup>, Charlotte Thålin<sup>6</sup>, Jochen M Schwenk<sup>1</sup>, Juan Carlos Souto Andres<sup>27</sup>, Pierre-  
15 Emmanuel Morange<sup>4,\*</sup>, Lynn Marie Butler<sup>1,3,5,7</sup> \*, David-Alexandre Tréguoët<sup>8,9,\*</sup>, Jacob  
16 Odeberg<sup>1,2,3,10,16,\*</sup>.

17 † Equal contribution

18 \*Equal contribution

19

20 **Author affiliations:**

- 21 1. Science for Life Laboratory, Department of Protein Science, CBH, KTH Royal Institute of  
22 Technology, SE-171 21 Stockholm, Sweden
- 23 2. Division of Internal Medicine, University Hospital of North Norway (UNN),  
24 PB100, 9038 Tromsø, Norway
- 25 3. Translational Vascular Research, Department of Clinical Medicine, UiT The Arctic  
26 University of Norway, Tromsø, Norway.

- 27 4. Aix-Marseille Univ, INSERM, INRAE, C2VN, Laboratory of Haematology, CRB  
28 Assistance Publique - Hôpitaux de Marseille, HemoVasc (CRB AP-HM HemoVasc),  
29 Marseille, France.
- 30 5. Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm,  
31 Sweden.
- 32 6. Division of Internal Medicine, Department of Clinical Sciences, Danderyd Hospital,  
33 Karolinska Institutet, Stockholm, Sweden.
- 34 7. Department of Clinical Chemistry, Karolinska University Hospital, Stockholm, Sweden.
- 35 8. University of Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR  
36 1219, team ELEANOR, Bordeaux, France.
- 37 9. Laboratory of Excellence GENMED (Medical Genomics), Bordeaux, France
- 38 10. Department of Medicine Solna, Karolinska Institutet and Karolinska University Hospital,  
39 Stockholm. Sweden.
- 40 11. Theme of Emergency and Reparative Medicine, Karolinska University Hospital,  
41 Stockholm, Sweden.
- 42 12. CEA, Centre National de Recherche en Génomique Humaine, Université Paris-Saclay,  
43 91057, Evry, France
- 44 13. Laboratory of Excellence GENMED (Medical Genomics), Evry, France.
- 45 14. Hematology Department and Biosurgical Research Lab (Carpentier Foundation),  
46 European Georges Pompidou Hospital, Assistance Publique Hôpitaux de Paris, 20 rue  
47 Leblanc, Paris, 75015, France
- 48 15. Innovative Therapies in Haemostasis, INSERM, Université de Paris, 4 avenue de  
49 l'Observatoire, Paris, 75270, France
- 50 16. Coagulation Unit, Department of Haematology, Karolinska University Hospital, SE-171  
51 76 Stockholm, Sweden

- 52 17. Department of Clinical Science, Intervention and Technology, Karolinska Institutet (KI),  
53 171 77, Stockholm, Sweden
- 54 18. Institute for Clinical Chemistry and Laboratory Medicine, University Medical Centre  
55 Hamburg-Eppendorf, D-20246 Hamburg, Germany
- 56 19. Center for Thrombosis and Hemostasis (CTH), Johannes Gutenberg University Medical  
57 Center, D-55131 Mainz, Germany
- 58 20. Irish Centre for Vascular Biology, School of Pharmacy and Biomolecular Sciences,  
59 Royal College of Surgeons in Ireland, Dublin 2, D02 YN77, Ireland
- 60 21. Genomics of Complex Diseases Group. Research Institute Hospital de la Santa Creu i  
61 Sant Pau. IIB Sant Pau, Barcelona, Spain
- 62 22. Department of vascular medicine, Paris Saint-Joseph Hospital Group, INSERM 1153-  
63 CRESS, University of Paris Cité, 185 rue Raymond Losserand, Paris, 75674, France
- 64 23. Laboratory of Excellence GENMED (Medical Genomics), Evry, France.
- 65 24. CEA, Centre National de Recherche en Génomique Humaine, Université Paris-Saclay,  
66 91057, Evry, France.
- 67 25. Centre D' Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris, France
- 68 26. Center for Physiology and Pharmacology, Institute of Vascular Biology and Thrombosis  
69 Research, Medical University of Vienna
- 70 27. Unitat d'Hemostàsia i Trombosi. Hospital de la Santa Creu i Sant Pau and IIB-Sant Pau,  
71 Barcelona, Spain.

72

73 **Address correspondence to:**

74 Jacob Odeberg

75 Science for Life Laboratory, Department of Protein Science, CBH, KTH Royal Institute of  
76 Technology, SE-171 21 Stockholm, Sweden

77 [Jacob.odeberg@scilifelab.se](mailto:Jacob.odeberg@scilifelab.se)

78 **or**

79 David-Alexandre Trégouët

80 University of Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR 1219, team

81 ELEANOR, Bordeaux, France.david-alexandre.tregouet@u-bordeaux.fr

82

83 **ABSTRACT**

84 Venous thromboembolism (VTE), comprising both deep vein thrombosis (DVT) and pulmonary  
85 embolism (PE) is a common, multi-causal disease with potentially serious short- and long-term  
86 complications. In clinical practice, there is a need for improved plasma biomarker-based tools for  
87 VTE diagnosis and risk prediction. We used multiplex proteomics profiling to screen plasma from  
88 patients with suspected acute VTE, and a case-control study of patients followed up after ending  
89 anticoagulant treatment for a first VTE. With replication in 5 independent studies, together totalling  
90 1137 patients and 1272 controls, we identify Complement Factor H Related Protein (CFHR5), a  
91 regulator of the alternative pathway of complement activation, as a novel VTE associated plasma  
92 biomarker. Using GWAS analysis of 2967 individuals we identified a genome-wide significant  
93 pQTL signal on chr1q31.3 associated with CFHR5 levels. We showed that higher CFHR5 levels  
94 are associated with increased thrombin generation in patient plasma and that recombinant  
95 CFHR5 enhances platelet activation *in vitro*. Thrombotic complications are a frequent feature of  
96 COVID-19; in hospitalised patients we found CFHR5 levels at baseline were associated with  
97 short-time prognosis of disease severity, defined as maximum level of respiratory support needed  
98 during hospital stay. Our results indicate a clinically important role for regulation of the alternative  
99 pathway of complement activation in the pathogenesis of VTE and pulmonary complications in  
100 acute COVID-19. Thus, CFHR5 is a potential diagnostic and/or risk predictive plasma biomarker  
101 reflecting underlying pathology in VTE and acute COVID-19.

102

## 103 **INTRODUCTION**

104 Venous thromboembolism (VTE), comprising both pulmonary embolism (PE) and deep vein  
105 thrombosis (DVT) is a common, multi-causal disease with serious short and long-term  
106 complications. VTE has a high mortality rate in the first year, especially within the first 30 days  
107 (~30% for PE) and a high risk of recurrence, with a cumulative incidence rate of 25% within 10  
108 years [1-4]. VTE diagnosis is challenging, as the predisposing common risk factors and clinical  
109 presentation can be consistent with multiple other conditions, particularly in the case of PE.  
110 Current VTE diagnostic work-up includes assessment of clinical probability, using clinical decision  
111 rules, e.g., the Well's score, in combination with elevated plasma D-dimer levels [5, 6]. However,  
112 D-dimer, a clot breakdown product, can be elevated in several other non-VTE conditions (e.g.,  
113 inflammation, surgery, cancer). Thus, its usefulness is limited to ruling out VTE in low probability  
114 cases. In medium and high probability cases, diagnostic imaging, i.e., compression ultrasound  
115 (CUS) on suspicion of DVT, or computed tomography pulmonary angiogram (CTPA) on suspicion  
116 of PE, is necessary to exclude or confirm diagnosis. These tests may not be readily available in  
117 the acute setting, resulting in delayed or incorrect diagnosis. Today, less than 20% of CTPAs and  
118 CUS confirm suspected PE [7-9] or DVT [10], respectively. Specific biomarker-based tools in the  
119 diagnostic workup could reduce unnecessary imaging, save resources and, in the case of CTPA,  
120 reduce exposure to radiation and contrast agent. Several studies have proposed VTE biomarker  
121 candidates (e.g., p-selectin, microvesicles) [11], but none have reached clinical implementation.  
122 There is also a need for improved tools and plasma biomarkers for risk prediction in VTE. Risk  
123 scores based on clinical risk factors and D-dimer levels have been developed for recurrence  
124 prediction [12-18], but none are routinely integrated into clinical practice. In addition to the genetic  
125 variants currently used in clinical assessment of hereditary thrombophilia (e.g., Factor V Leiden,  
126 prothrombin mutation) there are other more recently discovered common gene variants that  
127 contribute to VTE risk [19-21]. However, even when these are also incorporated into risk scores,  
128 they still lack sufficient precision for VTE prediction on an individual basis [22, 23]. This likely

129 reflects the interplay between transient and sustained risk factors in disease development,  
130 including acquired risk factors, genetics, and environmental exposures [24, 25].

131 VTE is a disease of the intravascular compartment and thus analysis of the blood proteome has  
132 the potential to capture the resulting effects of combined genetic, epi-genetic, and environmental  
133 contributors to risk variation. Large scale plasma proteomics screening allows for discovery of  
134 novel protein biomarkers with potential clinical utility for diagnosis and/or prediction. Such studies  
135 could also reveal biological pathways involved in VTE pathogenesis, for further functional studies  
136 to identify targets for tailored treatment. So far, a handful of plasma proteomics studies of VTE  
137 have been reported, presenting novel candidates associated with increased risk of VTE [26-30],  
138 including our own; the first affinity proteomics case-control study of patients followed up post  
139 treatment after a first VTE, reporting PDGFB as a novel VTE associated biomarker [27]. High  
140 throughput affinity proteomics studies have increased the capacity for discovery screening and  
141 identification of novel associations, as now it is possible to screen for hundreds or thousands of  
142 proteins in small plasma volumes [31]. Orthogonal verification is essential to confirm results  
143 generated using these technologies but is frequently absent from such studies.

144 Here, we aimed to identify novel biomarkers associated with acute VTE with a potential link to  
145 underlying VTE pathogenesis. We identified complement factor H-related protein 5 (CFHR5), a  
146 regulator of the alternative pathway of complement activation, as a novel VTE associated plasma  
147 biomarker. Furthermore, we found that CFHR5 plasma levels are associated with short term  
148 prognosis in acute COVID-19, a disease where thrombosis formation is central to pathology. Our  
149 study suggests that CFHR5 could be involved in the underlying pathogenesis of VTE and acute  
150 COVID-19, and that it is a potential clinical biomarker for thrombotic disease diagnosis and/or risk  
151 prediction.

152

153

154 **RESULTS**

155 **Affinity plasma proteomics identifies candidate proteins associated with acute VTE**

156 To identify plasma biomarkers for VTE we analysed samples collected as part of our *Venous*  
157 *thromboEmbolism BIOMarker Study (VEBIOS)* [27]. *VEBIOS* comprises two study arms; a  
158 prospective cohort of patients sampled at the Emergency Room (ER), Karolinska University  
159 Hospital, Sweden (*VEBIOS ER*) and a case/control study with patients sampled at an outpatient  
160 coagulation clinic after discontinuation of anticoagulant treatment after a first VTE (*VEBIOS*  
161 *Coagulation*). The discovery cohort, *VEBIOS ER* (Figure 1A), consisted of patients (n= 147)  
162 admitted to the ER with the suspicion of DVT in the lower limbs and/or PE. Following admission,  
163 two whole blood samples were collected from participants, into citrate and EDTA anticoagulant.  
164 Patient samples were classified as controls (n=96), when a VTE diagnosis was excluded by  
165 diagnostic imaging, and/or Well's clinical criteria with a normal D-dimer test, or cases (n=51) when  
166 VTE was confirmed by diagnostic imaging and anticoagulant treatment was initiated. A nested  
167 case/control sample set of 48 cases and 48 matched controls were selected for plasma protein  
168 analysis (Table 1). Target candidates for measurement were selected as previously described  
169 [27], based on: (i) indications from the literature, in house data or public repositories of a probable  
170 or plausible link to arterial or venous thrombosis (e.g., prior evidence of association with  
171 thrombosis or intermediate traits, or known involvement in biological pathways of relevance),  
172 including 124 that we predicted to have endothelial enriched expression [32], and (iii) the  
173 availability of target specific antibodies in the Human Protein Atlas (HPA). A total of 758 HPA  
174 antibodies, targeting 408 candidate proteins, were selected for incorporation into a single-binder  
175 suspension bead array (Figure S1 A and Table S1, Tab\_1), which was used to analyse plasma  
176 generated from the blood samples collected into citrate anticoagulant. The signal generated by  
177 antibody HPA059937, raised against the protein target sulfatase 1 (SULF1), was most strongly  
178 associated with VTE ( $p < 8.34E-06$ ) (Figure 1B, green point), with higher relative plasma levels in  
179 cases vs. controls (Figure 1D.i). Signals generated by a further seven antibodies were also



180 associated with VTE ( $p < 0.01$ ) (Table S1, Tab\_1). Protein signatures in plasma can be differently  
181 affected by the sample matrix; anticoagulants can inhibit specific proteases, influence soluble  
182 protein interactions, and modify analyte stability. Thus, the anticoagulant type used has potential  
183 consequences for biomarker identification [33]. We therefore replicated the *VEBIOS ER* discovery  
184 screen in EDTA samples drawn in parallel from the same patients (Figure 1C). Of the eight  
185 antibodies that produced signals associated ( $p < 0.01$ ) with VTE in the citrate samples (Table S1,  
186 Tab\_1), four were replicated in the EDTA samples (Figure 1B and C): HPA059937 (predicted  
187 target SULF1, green point), HPA044659 (predicted target Leukocyte surface antigen CD47  
188 [CD47], blue point), HPA003042 (predicted target Adenosine receptor A2a [ADORA2A], orange  
189 point) and HPA002655 (predicted target P-Selectin [SELP], red point) (Table S1, Tab\_1). In both  
190 anticoagulants, all four target candidates were elevated in cases vs. controls (Figure 1D and E, i-  
191 iv). In all subsequent experiments, citrated blood was used in the analysis, on the basis that it is  
192 the more commonly used sample anticoagulant in the clinical setting for coagulation analyses.

193 Previously, in the *VEBIOS Coagulation* study ( $n=177$ ) [27], we identified 29 protein candidates in  
194 plasma that were associated with prior VTE. This study was composed of patients sampled 1-6  
195 months after discontinuation of anticoagulation treatment (duration 6-12 months) following a first  
196 time VTE, or matched controls. Of the four antibodies that generated signals associated with  
197 acute VTE in citrate *and* EDTA plasma in *VEBIOS ER* (Figure 1B and C, marked with coloured  
198 circles), only HPA059937 (predicted target SULF1), produced a signal associated with prior VTE  
199 in the *VEBIOS Coagulation* study [27]. As our aim was to identify biomarkers associated with  
200 acute VTE that were potentially linked to the underlying disease pathogenesis, we prioritised this  
201 candidate on the basis that higher plasma concentrations observed in individuals with a  
202 documented increased risk of VTE (e.g., previous VTE) in *VEBIOS Coagulation* study indicated  
203 that it could also represent a constitutive and/or persistent risk factor.

204 **Complement Factor H Related protein 5 (CFRH5) is associated with VTE**

205 The antibodies used in the single-binder suspension bead arrays passed quality control for  
206 antigen binding specificity (see [www.proteinatlas.org/](http://www.proteinatlas.org/)), but selective binding to the target protein  
207 in context of the complex matrix of plasma requires verification, as antibody specificity and  
208 reliability can be a problematic issue [34-36] (Figure S1 C-F). To verify which protein(s) were  
209 captured by HPA059937 (predicted target SULF1) we performed immunocapture-mass  
210 spectrometry (IC-MS). Two proteins were bound to HPA059937 with a z-score>3 in triplicate  
211 experiments; the predicted target, SULF1 (z-score = 4.02, with 1 Peptide Spectrum Match [PSM]),  
212 and Complement Factor H Related protein 5 (CFHR5) (z-score=5.09, with  $\geq 21$  PSM) (Figure 1F  
213 and Table S1, Tab\_2). These data indicate that CFHR5 was the predominant protein captured by  
214 HPA059937 in plasma, whilst not ruling out concurrent binding of SULF1. High levels of CFHR5  
215 have been detected in plasma by mass spectrometry (MS) [37], but SULF1 (a protein predicted  
216 to be localised on the cell surface) is below the MS detection threshold [38]. To further verify  
217 CFHR5 binding specificity of HPA059937, we developed three dual binder assays (Figure S1B  
218 and Figure 1G), all with a commercial monoclonal antibody against CFHR5 as a detection  
219 antibody, combined with either: original antibody HPA059937 (Figure 1G.i), or one of two  
220 independent antibodies raised in house against CFHR5; HPA073894 (Figure 1G.ii) or  
221 HPA072446 (Figure 1G.iii), as bead coupled capture antibodies. When used to re-analyse the  
222 *VEBIOS ER* samples; all three assays consistently detected a higher level of target protein in  
223 cases vs. controls (Figure 1G iv-vi,  $p=0.0001$ ,  $0.0021$ ,  $0.0006$ , respectively). Furthermore, MFI  
224 values from all three strongly correlated with those generated by HPA059937 in the *VEBIOS ER*  
225 discovery screen (Figure 1G.vii-ix) (Spearman correlation 0.86, 0.78 and 0.84, respectively [all  
226  $p<0.00001$ ]). We made five dual binder assays that targeted SULF1, but none gave a quantitative  
227 signal in a plasma dilution series, or buffer containing a dilution series of recombinant SULF1  
228 protein. These data are consistent with CFHR5, as opposed to SULF1, being the target protein  
229 associated with VTE in the *VEBIOS ER* discovery screen.

230 **CFHR5 is associated with VTE independent of D-dimer or CRP**

231 The CFHR5 dual binder assay using capture antibody HPA072446 (Figure 1G.iii), together with  
232 recombinant CFHR5 protein standard, was used for absolute quantification of CFHR5 in *VEBIOS*  
233 *ER* and an extended sample set of the *VEBIOS Coagulation* study (n=284) (Table S2, Tab\_1 for  
234 cohort descriptive data). In the *VEBIOS ER* and *Coagulation* study, mean CFHR5 concentrations  
235 in control plasma samples were  $2842 \pm 756$  and  $2467 \pm 523$ ng/ml, respectively; levels in the range  
236 previously estimated by mass spectrometry ( $\sim 1900$ ng/ml) [37]. In the *VEBIOS ER* discovery  
237 study, the mean absolute CFHR5 concentration was confirmed as higher in patients with  
238 confirmed VTE, compared to patients where VTE was ruled out ( $3428$  ng/ml  $\pm 774$  [cases] vs.  
239  $2842$  ng/ml  $\pm 756$  [controls],  $p=1.05E-03$  [age and sex adjusted]); the odds ratio (OR) for diagnosis  
240 of acute VTE associated with one standard deviation (SD) increase of CFHR5 concentration was  
241  $2.54$  [confidence interval (CI)  $1.52-4.66$ ],  $p=1.05E-03$  (Figure 1H.i). Consistent with the relative  
242 quantification results in our previous study [27], absolute CFHR5 concentration was associated  
243 with prior diagnosis of VTE in the extended *VEBIOS Coagulation* cohort, compared to controls  
244 (mean concentration  $2681$  ng/ml  $\pm 554$  [cases] vs.  $2467$  ng/ml  $\pm 523$  [controls],  $p=8.42E-04$  [age  
245 and sex adjusted]); the OR for first time VTE was  $1.55$  [CI  $1.20-2.01$ ],  $p=8.85E-04$ ) (Figure 1J.i).  
246 We next investigated if CFHR5 levels were associated with VTE associated risk factors, such as  
247 age, body mass index [BMI]) and routine clinical laboratory tests for blood markers associated  
248 with thrombosis risk (e.g., D-dimer, c-reactive protein [CRP], thrombocyte count). In *VEBIOS ER*,  
249 CRP levels correlated with plasma CFHR5 concentration, in cases ( $\rho=0.57$ ,  $p<0.001$ ) and controls  
250 ( $\rho=0.52$ ,  $p<0.001$ ) (Figure 1H.ii), but there was no strong correlation between CFHR5 and the  
251 other parameters measured, in cases or controls, including D-dimer (Figure 1H.iii and Table S1,  
252 Tab\_3, Table A). In the *VEBIOS Coagulation* study, CFHR5 levels in cases correlated with both  
253 CRP ( $\rho=0.49$ ,  $p<0.001$ ) and D-dimer ( $\rho=0.42$ ,  $p<0.0001$ ) (Figure 1J.ii and iii, right panel), but in  
254 controls these correlations were weak (CRP  $\rho=0.24$ ,  $p=0.005$ ) or absent, respectively (Figure  
255 1J.ii, and iii, left panel and Table S1, Tab\_3, Table B). The association between CFHR5 and VTE

256 remained significant in both *VEBIOS ER* ( $p=0.0029$ ) and *VEBIOS Coagulation* ( $p=0.0015$ ) when  
257 adjusted for CRP. Therefore, we did not adjust for CRP in further analyses.

### 258 **CFHR5 is specifically expressed in liver hepatocytes with other complement related genes**

259 To further understand the expression characteristics of *CFHR5*, and to identify possible co-  
260 expressed or co-regulated proteins we used a whole transcriptome analysis approach. In a  
261 consensus dataset, consisting of normalized mRNA transcript (nTPM) levels across 55 different  
262 human tissue types, generated from Human protein Atlas (HPA) ([www.proteinatlas.org](http://www.proteinatlas.org)) [39] and  
263 Genotype-Tissue Expression Project (GTEx) ([www.gtexportal.org](http://www.gtexportal.org)) [40] datasets, *CFHR5* was  
264 highly and specifically expressed in the liver (Figure 2A). Single cell analysis of liver tissue [41],  
265 showed that *CFHR5* was specifically expressed in the hepatocyte cellular compartment (Figure  
266 2B). To identify transcripts potentially co-expressed or co-regulated with *CFHR5* in the liver, we  
267 analysed bulk RNAseq data ( $n=226$ ) retrieved from GTEx portal V8. We generated Pearson  
268 correlation coefficients between *CFHR5* and all other expressed protein coding transcripts  
269 (~19,525) (Table S1, Tab\_4, Table A) and used gene ontology (GO) and reactome analysis [42,  
270 43] to identify over-represented classes and pathways in the top 50 most highly correlated genes  
271 (Table S1, Tab\_4, Table B and C). Results were consistent with known *CFHR5* function;  
272 significant GO terms included '*complement activation*' (False Discovery Rate, FDR = $2.4E-16$ )  
273 and '*humoral immune response*' (FDR = $1.2E-12$ ) and reactome pathways included '*regulation of*  
274 *complement cascade*' (FDR= $3.8E-24$ ). We performed an unbiased weighted network correlation  
275 analysis (WGCNA) [44] on the same dataset, where correlation coefficients between all  
276 transcripts, excluding those classified as non-coding, were calculated and subsequently clustered  
277 into related groups, based on expression similarity (Table S1, Tab\_4, Table A [column F]).  
278 Transcripts that fulfilled the criteria of: (i) Pearson correlation with *CFHR5*  $>0.65$  and (ii) annotation  
279 to the *CFHR5*-containing gene cluster in WGNCA (group 68), were identified as those most likely  
280 to be co-expressed or co-regulated. Of these, 13/18 [72%] were other members of the  
281 complement cascade and 15/18 [83%] were also specifically expressed in liver [39] (Table S1,

282 Tab\_4, Table A, column F) (Table S1, Tab\_4, Table A, column B, bold text). When these proteins  
283 were interrogated using the protein-protein interaction database STRING, v11 [45], 13/18 had  
284 high confidence functional and physical associations (Figure 2C) (Table S1, Tab\_5, Table A).  
285 CFHR5 was most strongly linked to complement 3 (C3), the central hub of the largest of the three  
286 linked cluster groups identified (Figure 2C, clusters represented by green, red and cyan) (Table  
287 S1, Tab\_5, Table B).

### 288 **CFHR5 is associated with VTE independent of C3**

289 Plasma levels of C3 have previously been reported as associated with incident VTE [46]. To  
290 determine if the association between CFHR5 and VTE we observed is dependent or independent  
291 of the relationship with C3, we developed an in-house dual binder quantitative assay to measure  
292 C3 in the *VEBIOS ER* and *VEBIOS Coagulation* cohorts. In *VEBIOS ER*, plasma C3 was not  
293 elevated in cases, compared to controls (Figure 2D.i), CFHR5 and C3 did not significantly  
294 correlate in either group (Figure 2D.ii) and C3 was not associated with VTE (OR 1.04 [CI 0.68-  
295 158],  $p=0.86$ ). Conversely, in *VEBIOS Coagulation* C3 levels were higher in plasma from cases,  
296 compared to controls (Figure 1E.i), and CFHR5 and C3 correlated with each other in both  
297 ([controls  $\rho = 0.46$   $p<0.0001$ ], [cases  $\rho = 0.47$   $p<0.0001$ ]). After adjusting for age and sex, one  
298 SD increase in C3 level was significantly associated with VTE (OR 1.53 [CI 1.18-2.01],  $p=1.93E-$   
299 03). When adjusting for C3 together with age and sex, one SD increase in CFHR5 level was still  
300 associated with VTE in both *VEBIOS ER* (OR 2.65 [CI 1.53-5.01],  $p=0.00126$ ) and *VEBIOS*  
301 *Coagulation* (OR 1.37 [CI 1.03-1.83],  $p=0.032$ ). In contrast, when adjusting for CFHR5 together  
302 with age and sex, C3 was no longer significantly associated with VTE in *VEBIOS Coagulation*  
303 (OR 1.31 [CI 0.99-1.78],  $p=0.0645$ ). Thus, CFHR5 is associated with both a diagnosis of acute  
304 VTE and prior VTE, independent of C3.

### 305 **The CFHR5 association with VTE replicates in additional cohorts**

306 The identification of biomarkers associated with VTE diagnosis, or risk profiling, requires  
307 replication in independent cohorts, from different settings with different demographic profiles, to

308 determine feasibility for potential translation to clinical practice. We sourced three independent  
309 replication cohorts to test the association of CFHR5 with VTE; the Swedish Karolinska Age  
310 Adjusted D-Dimer study (*DFW-VTE*) VTE study (n=200) consisting of patients sampled at  
311 presentation with suspected VTE at Huddinge hospital, Stockholm that was either subsequently  
312 confirmed (cases; n=54) or ruled out (controls; n=146) [47] (Figure 3B), the French *FARIVE*  
313 case/control study (n=1158) consisting of patients sampled during the week following a diagnosis  
314 of acute VTE (n=582), with hospital-based controls (n=576) [48] (Figure 3C) and the Spanish  
315 *Riesgo de Enfermedad TROmboembólica VEEnosa (RETROVE)* study (n=668) of patients  
316 sampled post anticoagulant treatment (n=308), with population based controls (n=360) [49]  
317 (Figure 3E) (for all cohort details see Table S2, Tabs 2-4). The relative risk of VTE associated  
318 with CFHR5 per 1 SD increase in CFHR5 concentration was significant in all 3 replication cohorts:  
319 *DFW-VTE* (OR 1.80 [CI 1.29-2.58], p=0.0008) (Figure 3B), *FARIVE* (OR 1.24 [CI 1.10-1.40],  
320 p=0.0004) (Figure C), *RETROVE* (OR 1.29 [CI 1.09-1.53], p=0.002) (Figure 3E) (Table 2). In all  
321 cohorts, including *VEBIOS ER* and *VEBIOS Coagulation*, association results were comparable in  
322 sub-set analysis by thrombosis type (Table S1, Tab\_6, Table A) or sex (Table S1, Tab\_6, Table  
323 B). When samples from cases and controls were stratified according to CFHR5 concentration,  
324 the association with VTE risk was most pronounced in the third tertile, in all 5 cohorts analysed  
325 (Table 2).

### 326 **CFHR5 and risk of recurrent VTE**

327 The MARseille THrombosis Association Study (MARTHA) is a hospital-based cohort of over  
328 1,500 unrelated individuals recruited at the Thrombophilia center of La Timone hospital (Marseille,  
329 France). All patients had a history of a first VTE documented by venography, Doppler ultrasound,  
330 angiography and/or ventilation/perfusion lung scan [50]. We measured plasma CFHR5  
331 concentrations in samples from VTE patients where follow up data was available (n=669) (for  
332 cohort details see Table S2, Tab\_5). Using a Cox survival model with left truncature on the VTE  
333 cases we tested for association between CFHR5 and recurrence. In total, there were 124

334 recurrent events (52 among 231 males, 72 among 438 females). After adjusting for sex, familial  
335 history of VTE, provoked or unprovoked status of the first VTE, age at first VTE, and BMI, the  
336 Hazard Ratio (HR) associated of 1 SD increase in CFHR5 levels was HR = 1.13 [0.96-1.32], p =  
337 0.134. The association was consistent between females (HR = 1.11) and males (HR = 1.14) and  
338 between patients with DVT (HR = 1.18) or PE as first event (HR=1.13). This association was  
339 driven by the subgroup of patients with unprovoked first VTE (HR = 1.32 [0.99 – 1.77], p=0.056),  
340 as no association was observed when the first event was provoked (HR = 1.01 [0.83 – 1.23],  
341 p=0.900). However, the testhomogeneity between these two HRs did not reach 0.05 significance  
342 (p = 0.23).

### 343 **Genome Wide Association Study on CFHR5 plasma levels**

344 To explore if CFHR5 concentration in plasma was influenced by genetic variants, we performed  
345 a meta-analysis of GWAS data CFHR5 concentrations in individuals from the *FARIVE* (n=1,033),  
346 *RETROVE* (n=668) and *MARTHA* (n=1,266) studies. The results from the association results are  
347 summarised in Figure 4A. Of N= 7,135,343 SNPs tested in a total sample of 2967 individuals, one  
348 genome-wide significant (p < 5E-08) signal was observed on chr1q31.3. The lead SNP at this  
349 locus was rs10737681, mapping to CFHR1/CFHR4, and the G allele was associated with  $\beta =$   
350  $+81.33 \pm 8.67$  (p = 6.49E-21) increase in CFHR5 levels. After conditioning on the rs10737681, a  
351 borderline significant association (p = 9.83E-08) was observed at the rs10494747 where the A-  
352 allele tended to associate with an increase of  $\beta = +47.60 \pm 8.92$  in CFHR5 levels. Rs10494747 is  
353 distant of ~343 kb from CFHR2 rs10737681, maps to ZBTB41 (Figure 4B) and was in moderate  
354 linkage disequilibrium ( $r^2 \sim 0.10$ ,  $D' = -0.63$ ) with the lead rs10494747 variant. Together, these two  
355 variants only explained 2.2% of the inter-individual variability in CFHR5 plasma levels. Neither  
356 SNP was associated with VTE risk in the *FARIVE* and *RETROVE* case-control samples (Table  
357 S1, Tab\_7).

### 358 **CFHR5 enhances platelet activation and degranulation in plasma**

359 C3a, generated by cleavage of C3, can increase platelet activation [51-53]. As CFHR5 has a  
360 regulatory role upstream of C3/C3a activation, we investigated the effect of recombinant CFHR5  
361 on platelet activation *in vitro*. Platelet rich plasma was pre-incubated with 6µg/ml recombinant  
362 CFHR5, a concentration corresponding to the upper range of that detected in the plasma of the  
363 VTE case group in *VEBIOS ER*, and in agreement with what has been reported in previous  
364 literature [54]. Platelet activation was measured by surface expression of P-selectin, activated GP  
365 IIb/IIIa or CD63 (Figure 5A, B and C, respectively) in response to adenosine diphosphate (ADP),  
366 convulxin or TRAP6 (Figure 5 i, ii and iii, respectively). Under baseline conditions (in the absence  
367 of stimulation) activation was not modified by pre-incubation with CFHR5 (Figure 5A.i-iii).  
368 Following stimulation with ADP, a higher percentage of platelets pre-incubated with CFHR5  
369 expressed P-selectin (Figure 5A.i) ( $p=0.0056$ ), activated GP IIb/IIIa (Figure 5B.i) ( $p=0.031$ ) and  
370 CD63 (Figure 5C.i) ( $p=0.009$ ), compared to the control. Pre-incubation with CHFR5 also  
371 potentiated platelet activation in response to convulxin or TRAP6 stimulation (Figure 5A-C.ii and  
372 iii), although the effect appeared to be more strongly linked to stimulus concentration, than that  
373 observed for ADP. Washed platelet response to ADP, convulxin or TRAP6 was not modified by  
374 preincubation with CFHR5 (Figure S2) (ANOVA all  $p>0.05$ ), indicating that additional components  
375 in plasma are required for the observed effects, and that they are not a direct effect of CFHR5 on  
376 platelets. These data are consistent with the concept that CFHR5 is involved in the underlying  
377 pathology of VTE.

### 378 **CFHR5 is associated with thrombin generation potential in patients with previous VTE**

379 As thrombin generation has been associated with increased risk of VTE [55], we tested the  
380 association between CFHR5 plasma concentration and thrombin generation as measured by  
381 thrombinoscope in *MARTHA* ( $n=774$  VTE cases, see Table S2, Tab\_5 for details) with replication  
382 in *RETROVE* (308 cases/360 controls, see Table S2\_Tab\_4 for details). In both *MARTHA* and  
383 *RETROVE* cases, we find significant association between CFHR5 and lag time ( $\rho=0.181$ ,  $p=$   
384  $p<0.0001$  and  $\rho=0.176$ ,  $p<0.0001$ , respectively), Endogenous Thrombin Potential (ETP)



385 (rho=0.105, p= 0.0036, and rho=0.130, p<0.0001, respectively), peak (rho=0.117, p= 0.0012, and  
386 rho=0.132, p<0.0001), and ttPeak (rho=0.116, p=0.0012, and rho=0.086, p=0.0274 (see Table  
387 S1, Tab\_8).

### 388 **Plasma CFHR5 is associated with prognosis and respiratory status in COVID-19 patients**

389 Accumulating evidence shows that vascular dysfunction, a prothrombotic state and complement  
390 activation underlies severe COVID-19 pathophysiology, manifested as respiratory failure linked  
391 to microvascular thrombosis in lung [56, 57]. Furthermore, VTE is a frequent complication in  
392 hospitalized COVID-19 patients [58]. To determine if CFHR5 was associated with COVID-19  
393 severity, we measured relative plasma levels in 339 samples collected at consecutive timepoints  
394 from 112 hospitalised COVID-19 patients in the *COVID-19 biomarker and Immunity*  
395 (*COMMUNITY*) study [59] (Figure 6A) (cohort details see Table S2, Tab\_6), using a single binder  
396 assay (antibody HPA073894). At each sampling timepoint, we categorised patients into one of  
397 four ‘*Respiratory support Index*’ (RI) groups, as a proxy variable for COVID-19 severity; **RI 0**: no  
398 respiratory support required, **RI 1**: ≤ 5 Litres of oxygen on nasal cannula or mask, **RI 2**: > 5 Litres  
399 of oxygen on nasal cannula or mask, **RI 3**: non-invasive ventilation/high-flow nasal cannula, or **RI**  
400 **4**: intubation (Figure 6B). CFHR5 levels were positively associated with patient RI at the time of  
401 sample collection (p=<0.0001, ANOVA) (Figure 6C.i). To determine if the measurement of plasma  
402 CFHR5 at admission had potential value for prognosis, we performed a longitudinal analysis to  
403 determine the association with maximum respiratory support index (max RI) registered at *any*  
404 sampling time point during the hospital stay. CFHR5 at baseline was associated with the  
405 maximum level of respiratory support required (Figure 6C.ii) (p=<0.0022, ANOVA). Therefore,  
406 CFHR5 levels measured early upon admission carry short-term prognostic information of disease  
407 development.

408

## 409 **DISCUSSION**

410 Here, we aimed to identify biomarkers associated with acute VTE that are linked to disease  
411 pathogenesis. Using a nested case-control study, derived from a cohort of patients presenting to  
412 the ER with suspected acute VTE, and from a case-control study with patients that had suffered  
413 a previous first VTE, we identify CFHR5, a regulator of the alternative complement activation  
414 pathway, as such a biomarker. The association of CFHR5 with current or previous VTE was  
415 replicated in three additional cohorts or case-control studies, and we also found a trend for  
416 association with risk for recurrence of unprovoked VTE. We identify a genome-wide significant  
417 signal associated with CFHR5 levels located in the CFHR1-5 gene cluster loci. We further provide  
418 evidence of a direct role of CFHR5 in the induction of a pro-thrombotic phenotype, through its  
419 effect on platelet activation. Finally, we show that increased CFHR5 was associated with degree  
420 of respiratory insufficiency and prognosis in acute COVID-19. Our findings indicate CFHR5 has  
421 potential application as a clinical biomarker for VTE diagnosis and risk prediction, providing further  
422 support to the idea that complement regulation is a key element of VTE pathogenesis.

423 Currently, D-dimer is the only plasma biomarker used in VTE diagnostic work-up, but its clinical  
424 utility is limited to ruling out VTE in low-risk patients. Several studies have attempted to identify  
425 novel biomarkers with potential clinically usefulness for the confirmation of VTE diagnosis, and  
426 although a number have been identified [11], none have yet been implemented in clinical practice.  
427 For many, like D-dimer, elevated levels are a consequence of thrombosis formation, e.g.  
428 biomarkers of fibrinolysis, clot re-modelling or resolution (e.g. MMPs), inflammation secondary to  
429 local vascular and tissue injury (e.g., CRP, IL-6, IL-10, fibrinogen), or of endothelial and/or platelet  
430 activation (e.g. vWF, P-selectin) [11, 60, 61].

431 We find no correlation or association between D-dimer and CFHR5 in the acute VTE setting,  
432 supportive that increased concentration at diagnosis is not secondary to thrombus formation. In  
433 contrast, we found a strong correlation between D-dimer and CFHR5 levels in patients followed  
434 up after ending treatment for a first VTE, but not in controls. D-dimer has been associated with

435 increased risk of first and recurrent VTE [12-18, 62] and thus, our results are consistent with a  
436 link between CFHR5 and subclinical coagulability in these patients at follow-up, possibly due to  
437 persistent risk factors.

438 CFHR5 shares sequence and structural homology with Complement Factor H (CFH) [63], the  
439 main negative regulator of alternative pathway (AP) activation in plasma [64]. Under normal  
440 conditions, the AP is constitutively active through spontaneous hydrolysis of the thioester bond in  
441 C3 and the formation of the initial fluid phase C3 convertase, C3(H<sub>2</sub>O)Bb, which cleaves C3 into  
442 C3a and C3b [64]. CFH promotes decay of the alternative and classical pathway convertases and  
443 is a cofactor in the cleavage of C3b, hereby regulating excess activation of C3 [65]. CFH inhibits  
444 C3 convertase by competing with Complement Factor B (CFB) for binding to C3, preventing  
445 formation of C3a. CFHR5 antagonizes CFH function, through competitive binding to C3b and its  
446 fragment C3d [66], thus deregulating AP activation. CFHR5 also promotes complement activation  
447 by interfering with CFH binding to CRP, pentraxin 3 (PTX3), and extracellular matrix (ECM) [67].  
448 The complement and haemostatic systems interact at several points during initiation, propagation,  
449 and regulation of complement activation and coagulation [68]. Studies have indicated a role of  
450 complement in VTE pathogenesis [46, 52], but underlying mechanisms are not well understood.  
451 Elevated plasma C3 in baseline samples has been shown to be associated with increased risk of  
452 future VTE [46]. Consistent with these findings, C3 was associated with prior VTE in the *VEBIOS*  
453 *coagulation* study, but not with acute VTE in the *VEBIOS ER* study. In both cases, and in the  
454 previous study by Nordgaard *et. al.* [46], total C3 level, rather than the active form (C3a) is  
455 measured; it is possible that in acute VTE, regulation of C3 convertase (by CFHR5) is important,  
456 rather than absolute C3. It could be speculated that the association of C3 with VTE in individuals  
457 sampled pre-VTE [46] or following treatment, reflects co-regulation of CFHR5 and C3 under basal  
458 conditions, where C3 is a proxy for CFHR5, where CFHR5 represents the functional link to VTE  
459 risk, rather than C3 levels *per se*. Consistent with this idea, in the *VEBIOS coagulation* study, the  
460 association between CFHR5 and previous VTE remained significant when adjusted for C3.

461 The *CFHR5* locus maps to chromosome 1q31.3 at one end a gene cluster that spans  
462 approximately 350 kb and that includes (in order from *CFHR5*) the *CFHR2*, *CFHR4*, *CFHR1*,  
463 *CFHR3*, and *CFH* loci (Figure 4B). The rs10737681 we identify with genome wide association  
464 with *CFHR5* levels maps between the *CFHR4* and *CFHR1* genes (Figure 4B). Of note,  
465 the *CFHR2* locus has just been identified as a novel susceptibility locus for VTE in the recently  
466 published international effort on VTE genetics [21]. The lead SNP at this locus is rs143410348,  
467 which is in moderate LD with the rs10737681 ( $r^2 = 0.38$  in *FARIVE*) that here we found associated  
468 with *CFHR5* plasma levels. Unfortunately, rs143410348 was not imputed in *MARTHA* nor in  
469 *RETROVE*, preventing us from investigating its association with *CFHR5* plasma levels in our  
470 meta-analysis. Nevertheless, in *FARIVE* where it was imputed, its association with *CFHR5* levels  
471 tended to be slightly less pronounced than that observed with rs10737681 ( $+76.4 \pm 27.0$ ;  $p =$   
472  $4.8E-03$  vs  $+101.5 \pm 29.1$ ;  $p = 5.06E-04$ ). Altogether, these observations emphasize the need for  
473 a deeper investigation of the genetic architecture of the *CFHR1/CFHR4/CFHR2/CFHR5*  
474 locus with respect to VTE risk.

475 Our study indicates that *CFHR5* has a possible role in the potentiation of platelet activation. Using  
476 a flow restricted tissue-factor dependent mouse model of VTE, Subramaniam *et. al.* showed that  
477 C3 and C5 affected platelet activation and tissue factor procoagulant activity by different  
478 mechanisms, independent of formation of the terminal complement C5b-C9 complex [53]. In  
479 ligated inferior vena cava, both C3 deficient and C5 deficient mice developed smaller thrombi,  
480 with reduced fibrin deposition compared to the wild type. C3, but not C5, deficient mice had  
481 reduced platelet activation *ex vivo*, reduced platelet deposition *in vivo*, and reduced thrombosis  
482 incidence (<30 vs. 80% in wild type). These data indicate that in VTE C3 has an important role in  
483 initial hemostasis, independent of downstream complement proteins [52]. C3a, acting through  
484 platelet receptor C3aR, has an important role in the activation of the Gp3a/2b fibrinogen receptor  
485 via intraplatelet signaling, and subsequent thrombosis formation [51]. Similar to the study of  
486 Subraminam *et. al.* [53], Sauter *et. al.* showed C3 deficient mice had prolonged bleeding time,

487 that could be reversed by intravenous administration of C3a peptide [51]. The C3a-C3aR induced  
488 intracellular signaling was mediated through the Rap1b activation, where co-stimulation of  
489 platelets with C3a-ADP resulted in increased Rap1b activity on top of the platelet stimulation by  
490 only ADP. In our study, we observe a similar co-stimulatory effect of CFHR5 on ADP- (and  
491 convulxin- or TRAP6-) induced platelet activation. This effect was observed on platelets in  
492 plasma, but not on those that were pre-washed, consistent with the effect of CFHR5 on platelet  
493 activation being due to its interaction with other complement factors (i.e., C3) in plasma. On basis  
494 of these recently published mechanistic findings, our *in vitro* results indicate that CFHR5  
495 regulation of the alternative pathway of activation has an important role in C3a mediated platelet  
496 activation in thrombosis, providing a potential functional link to its association with acute VTE.  
497 Our findings indicate that CFHR5 mediated dysregulation of the AP could play an important role  
498 in COVID-19 progression. Acute respiratory distress syndrome (ARDS) is the most common  
499 serious complication of COVID-19 [69], where patients suffer from dysregulated haemostasis,  
500 presenting with atypical ARDS [69, 70] with thrombosis formation in the lung microvasculature  
501 [71]. Other studies have shown that severe disease is associated with complement activation [56],  
502 and the cross-talk between coagulation and complement contributes to elevated coagulopathy  
503 and thromboembolic complications [72]. Furthermore, elevated complement activation has been  
504 correlated with disease severity and development of ARDS in hospitalised COVID-19 patients  
505 [72] and patients with severe COVID-19 have a higher incidence of VTE [73].  
506 Our study has various strength and limitations; *VEBIOS ER*, the discovery cohort, that was  
507 derived from a single centre, where blood sampling for plasma biobanking was performed in  
508 parallel to that for routine tests after initial evaluation (before diagnostic imaging or anticoagulant  
509 treatment), thus avoiding bias in inclusion or biobanking. Samples were handled according to  
510 standard clinical chemistry lab routine, thus variations in needle-to-spin-to-freeze time were  
511 equivalent between case and control samples. As biobanking was based on the routine sample  
512 flow, this increases the feasibility that identified biomarker candidates are suitable for clinical

513 translation into a routine setting. Importantly, we demonstrate an association of CFHR5 with VTE  
514 in several independent studies, that include patients in the acute setting, at follow up, and prior to  
515 recurrence. One limitation of our study is that we have not analysed a cohort of individuals that  
516 were sampled prior to VTE event. From a technological perspective, our study demonstrates the  
517 need for orthogonal verification of any potential biomarker identified using antibody-based  
518 proteomics screening [34, 35]. The same caution should be extended to findings generated using  
519 other high throughput affinity proteomics technologies vulnerable to non-specific protein binding,  
520 such as aptamer-based [74], where missense single nucleotide polymorphisms can affect binding  
521 in a manner where a genetic difference drive associations, rather than protein levels (Figure S1  
522 F) [75, 76]. Studies comparing different affinity proteomics technologies have found correlations  
523 of proteins assayed with two or more platforms to range from highly concordant ( $r = 0.95$ ) to  
524 inversely correlated ( $r = -0.48$ ) [75], highlighting further the need for orthogonal validation of any  
525 potential biomarker identified.

526 The next steps towards the translation of our findings into a clinical setting is to develop and  
527 establish standardised methods for quantification, to establish reference intervals and define cut  
528 off values with respect to specificity and sensitivity. Furthermore, it remains to be established if  
529 the incorporation of CFHR5 measurements into clinical decision rules or other scores can improve  
530 predictive power. The inclusion of CFHR5 measurements as a diagnostic and/or risk predictive  
531 marker in randomized clinical trials of acute VTE and VTE recurrence would be particularly  
532 informative, as these are two areas of high clinical relevance in need of improved tools for clinical  
533 decision making. Furthermore, while our study indicates that CFHR5 has a functional role in VTE  
534 development, further studies are needed to understand the mechanism.

## 535 MATERIALS AND METHODS

### 536 PATIENTS AND SAMPLES

#### 537 Discovery study

##### 538 *Venous thromboembolism biomarker study (VEBIOS)*

539 VEBIOS is part of a collaboration between Karolinska University Hospital, Karolinska Institute and  
540 Royal Institute of Technology (KTH) designed to identify new plasma biomarkers for VTE [27].

541 VEBIOS comprises two different studies: (i) *VEBIOS ER study* is a prospective cohort study  
542 carried out at the Emergency Room (ER) at the Karolinska University Hospital in Solna, Sweden,

543 between December 2010 and September 2013. All patients admitted with the suspicion of deep

544 vein thrombosis (DVT) in the lower limbs and/or pulmonary embolism (PE), over 18 years old

545 were eligible for the study. Exclusion criteria were patients with on-going anticoagulant treatment,

546 pregnancy, active cancer, short life expectancy or lack of capacity to leave approved consent. A

547 case was defined if a) VTE was confirmed by diagnostic imaging - compression venous

548 ultrasonography (CUS) in patients with suspected DVT in the lower limbs, and computed

549 tomography pulmonary angiography (CTPA) in patients with suspected PE, and b) anticoagulant

550 treatment was initiated based on the VTE diagnosis. Patients with no evidence of an acute VTE,

551 (neither by diagnostic imaging nor by Well's clinical criteria) that had a normal d-dimer test [5],

552 were referred as controls in the study. All participants were sampled before any anticoagulant

553 treatment. Whole blood was collected in citrate or EDTA anticoagulant at the ER and sent within

554 30 minutes to the Karolinska University Laboratory. After centrifugation at 2000g for 15 minutes,

555 plasma aliquots were snap frozen and stored at -80°C. *Data collection:* For each patient, doctors

556 filled in a questionnaire detailing (1) any provoking factors within one month preceding the visit to

557 the ER (2) current health situation, alcohol consumption and smoking habits; (3) family history of

558 VTE (4) ongoing antithrombotic (antiplatelet) treatment and (5) estrogen containing

559 contraceptives and hormone replacement therapy (women only). Information from the ER visit on

560 patient weight and height (when available) along with results from routine laboratory tests i.e.,

561 blood count, d-dimer, C-reactive protein (CRP), creatinine, international normalized ratio (INR)  
562 and activated partial thromboplastin time (aPTT) were extracted from the medical records. In total,  
563 158 patients were included (52 cases). For the present study, 48 cases were available for analysis  
564 and 48 controls were matched as closely as possible by age and sex. Clinical characteristics of  
565 the sample set is given in Table 1. (ii) *VEBIOS Coagulation study* is an on-going case-control  
566 study established January 2011 of patients sampled at an outpatient coagulation clinic after  
567 discontinuation of anticoagulant treatment after a verified first VTE (DVT to the lower limbs and/or  
568 PE), matched with healthy controls from the population. Patients were between 18 to 70 years of  
569 age, free from cancer, severe thrombophilia and pregnancy at inclusion [27]. In the present study,  
570 we analysed an extended sample set of *VEBIOS Coagulation* comprising all available samples;  
571 144 cases and 140 controls (Table S2, Tab\_1). Approval for *VEBIOS* was granted by the regional  
572 research ethics committee in Stockholm, Sweden (KI 2010/636-31/4) and all participants gave  
573 informed written consent, in accordance with the Declaration of Helsinki.

#### 574 **Replication cohorts**

575 The Swedish Karolinska Age Adjusted D-Dimer study (*DFW-VTE study*) includes patients with  
576 clinically suspected acute VTE, prospectively recruited from the ER of Karolinska University  
577 Hospital in Huddinge, Stockholm, as previously described [47]. The patients were out-patients  
578 with low-to-high probability of acute PE or DVT in a lower limb. The study was approved by the  
579 regional ethics review board in Stockholm (DNR 2013-2143-31-2), and all participants gave  
580 informed written consent, in accordance with the Declaration of Helsinki. For the current study,  
581 biobanked plasma aliquots were available for a subset of subjects comprising 15 patients with  
582 PE, 39 with DVT, and 146 controls where VTE was excluded. Controls were identified based on  
583 negative diagnostic imaging, or a low Well's score together with negative D-dimer. Clinical  
584 characteristics are described in Table S2, Tab\_2.

585 The *FARIVE study* is a French multicentre case-control study carried out between 2003-2009, as  
586 previously described [77]. The study consists of patients with first confirmed VTE (DVT to the



587 lower limbs and/or PE) from 18 years of age, matched to hospital controls with no previous  
588 thrombotic event. All participants were free from cancer. The study was approved by the Paris  
589 Broussais-HEGP ethics committee in Paris (2002-034) and all participants gave informed written  
590 consent, in accordance with the Declaration of Helsinki. In the current study we used a subset of  
591 *FARIVE* samples (n=1158), as previously described [27] [78]. From most cases, blood was  
592 collected in the first week after diagnosis and during anticoagulant treatment initiation. Clinical  
593 characteristics are described in Table S2, Tab\_3.

594 The *Riesgo de Enfermedad TROmboembólica VEnosa (RETROVE)* study is a prospective  
595 case-control study of 400 consecutive patients with VTE (cancer associated thrombosis  
596 excluded) and 400 healthy control volunteers. Individuals were recruited at the Hospital de la  
597 Santa Creu i Sant Pau of Barcelona (Spain) between 2012 and 2016. Controls were selected  
598 according to the age and sex distribution of the Spanish population (2001 census) [79]. All  
599 individuals were  $\geq 18$  years. All procedures were approved by the Institutional Review Board of  
600 the Hospital de la Santa Creu i Sant Pau, and all participants gave informed written consent, in  
601 accordance with the Declaration of Helsinki. In the current study, samples from 308 cases and  
602 360 controls were used. Clinical characteristics are described in Table S2, Tab\_4.

603 The *Marseille Thrombosis Association study (MARTHA)* is a population based single centre study,  
604 as previously described [78]. Recruitment in *MARTHA* started in 1994 at Timone Hospital in  
605 Marseille (France) and is still ongoing. The cohort from 1994 and 2008, includes a total of 1542  
606 VTE-cases (66% women) that donated blood for further analysis. Ethical approval was granted  
607 from the Department of Health and Science, France (2008-880 & 09.576) and all participants  
608 gave informed written consent, in accordance with the Declaration of Helsinki. In the current study,  
609 proteomics data generated for 1322 sampled *MARTHA* cases was used. For 669 of the *MARTHA*  
610 cases, follow up data up to 12 years post-event was available and used to analyse risk of recurrent  
611 VTE. For a subset of 774 cases data for thrombin generation potential (TGP) was available [80],

612 which was used to analyse the association between CFHR5 and blood coagulability. Clinical  
613 characteristics are described in Table S2, Tab\_5.

614 The *COVID-19 biomarker and Immunity (COMMUNITY) study* is a single centre study of 112  
615 patients with COVID-19 admitted to a general ward, intermediate units, or intensive care units at  
616 Danderyd Hospital, Stockholm Sweden between April 15<sup>th</sup> and May 27<sup>th</sup>, 2020. Inclusion required  
617 confirmed SARS-CoV2 infection, based on viral RNA detection by reverse-transcriptase  
618 polymerase chain reaction of nasopharyngeal or oropharyngeal swabs, or clinical presentation of  
619 COVID-19. Exclusion criteria were age <18 years. Patients were followed longitudinally from  
620 inclusion, and blood samples were collected shortly after hospital admission and every 2-3 days  
621 during the hospital stay. Procedures for blood sampling and plasma preparation have been  
622 previously described [59]. Demographic data, routine lab results, comorbidity and information and  
623 variables, reflecting clinical deterioration, including respiratory support were obtained from  
624 medical records. Patients were divided into groups based on respiratory support required at the  
625 time of sampling ('Respiratory support Index' or RI): RI = 0; no respiratory support, RI = 1; ≤5L of  
626 oxygen on nasal cannula or mask, RI = 2; >5L of oxygen on nasal cannula or mask, RI = 3; non-  
627 invasive respiratory support and RI = 4; intubation. Level of respiratory support and oxygen  
628 supplementation were set at by the treating physician. For the current study, 339 samples  
629 collected from 112 patients were analysed, with at least 2 samples for 71 patients (63.4%). The  
630 study was approved by the National Ethical Board (EPM 2020-01653). Clinical characteristics are  
631 described in Table S2, Tab\_6.

## 632 **ANALYSIS OF PLASMA BY TARGETED PROTEOMICS**

633 Plasma proteomic profiles in *VEBIOS ER* were generated using multiplexed suspension bead  
634 arrays (SBA) with 758 antibodies selected from the Human Protein Atlas (HPA) project, targeting  
635 408 proteins (Table S1), using identical design, procedures and methods as previously described  
636 [27]. Briefly, paired samples were randomly distributed within the same 96-well area. Two  
637 suspension bead arrays composed of 380 antibodies and 4 controls were used to sequentially

638 generate profiles of the 96 samples in parallel. Proteomics profiling was performed in both EDTA  
639 and Citrate plasma. Median fluorescence intensity (MFI) values were obtained from the  
640 suspension bead array assay by detecting at least 32 beads per ID and sample with the FlexMap  
641 3D instrument (Luminex® Corp) and were normalized by (1) probabilistic quotient normalization  
642 as accounting for any potential sample dilution effects [81], and (2) multidimensional MA (M=log  
643 ratio; A=mean average, scales) normalization to minimize the difference amount the subgroups  
644 of the samples generated by experimental factor as multiple batches [82]. Association of target  
645 proteins with VTE was tested using linear regression analysis while adjusting for age and gender.  
646 Log-transformation was applied to reduce any skewness in the proteomic data distribution. The  
647 analyses were performed using the R statistical computing software [83].

#### 648 **Immunocapture mass spectrometry (IC-MS)**

649 IC-MS was performed in triplicate, as previously described [27] using the HPA059937 antibody  
650 (Atlas Antibodies) or rabbit immunoglobulin G (rlgG) as a negative control. Digested samples  
651 were analyzed using an Ultimate 3000 RSLC nanosystem (Dionex) coupled to a Q-Exactive HF  
652 (Thermo). Uniprot complete human proteome (update 20180131) was used to query the raw data,  
653 with the engine Sequest and Proteome Discoverer platform (PD, v1.4.0.339, Thermo Scientific).  
654 An internal database containing the most common proteins detected by IC-MS in plasma was  
655 used to calculate Z-scores [36].

#### 656 **In-house developed bead based dual binder immunoassays**

657 A Suspension Bead Array (SBA) was built with the capture antibodies raised against human  
658 extracellular sulfatase 1-SULF1 (rabbit polyclonal HPA059937) and human CFHR5 (rabbit  
659 polyclonal HPA072446 and HPA073894) covalently coupled to color-coded magnetic beads as  
660 previously described (Drobin et al., 2013; Neiman et al., 2013). Bead-coupled rabbit IgG (Bethyl  
661 Laboratories Inc.) and mouse-IgG and bare beads were included as negative controls. Mouse  
662 anti-human SULF1 antibodies Abnova ABIN525031, Abcam ab172404, Thermofisher PA5-  
663 113112, Human Protein Atlas antibodies HPA054728 and HPA051204, and mouse monoclonal

664 anti-human CFHR5 (R&D systems, MAB3845) antibody were labelled with biotin and used as  
665 detection antibodies in combination with their respective capture antibodies.—Citrate plasma  
666 samples were thawed on ice and centrifuged for 1 min at 2000 rpm and diluted in buffer polyvinyl  
667 casein 10% rlgG (PVXcas 10% rlgG; polyvinyl alcohol, Sigma Aldrich P8136;  
668 polyvinylpyrrolidone, Sigma Aldrich PVP360; Blocker Casein, Thermo 37528), heated at 56°C for  
669 30 min and incubated with the SBA overnight. The detection antibody was used at 1ug/mL for 90  
670 min, and streptavidin- R-phycoerythrin (R-PE) conjugate (Life Technologies; SA10044) was used  
671 for the fluorescence read out in Luminex platform. The dual binder assays based on HPA059937,  
672 HPA072446 or HPA073894 as capture antibodies together with the monoclonal anti-human  
673 CFHR5 were used to measure samples in the *VEBIOS ER* cohort, and correlations between MIF  
674 values obtained with the 3 different dual binder assays and the original single binder assay with  
675 HPA059937 were determined.

#### 676 **Absolute quantification of CFHR5, C3 and D-dimer**

677 For CFHR5 quantification, rabbit polyclonal anti-human CFHR5 HPA072446 (Atlas Antibodies)  
678 and mouse monoclonal anti-human CFHR5 (R&D systems; MAB3845) antibodies were used in a  
679 dual binder assay. Human recombinant CFHR5 (R&D systems; 3845-F5) spiked into chicken  
680 plasma (Sigma Aldrich, St. Louis, United States, P3266) was used as a standard. All samples  
681 were diluted 1:300 in PVXcas 10% rlgG. For C3 quantification, mouse anti-human C3 and mouse  
682 monoclonal anti-human C3 antibodies (Bsi0263, Bsi0190, respectively, Biosystems International)  
683 were used in a dual binder assay. Human recombinant C3 (Sigma Aldrich, C2910) spiked into C3  
684 depleted serum (Merck, 234403) was used as a standard. All samples were diluted 1:5000 and  
685 analysed as described above. D-dimer was quantified by ELISA (D-Di 96 test, product #00947,  
686 Asserachrom, France) following the manufacturer's instructions. In the *DFW-VTE* study, D-dimer  
687 values were analysed at the Karolinska University Hospital Laboratory in fresh samples sent for  
688 routine clinical chemistry analysis, as part of the work up in the ER.

#### 689 **Analysis of CFHR5 in COVID-19 patients in the *COMMUNITY* study**

690 The HPA antibody HPA073894, targeting the CFHR5 protein, was included in a screening panel  
691 using multiplexed suspension bead arrays (SBA) following a similar protein profiling protocol as  
692 in the targeted plasma proteomics screening (see above). Association of CFHR5 plasma levels  
693 with Respiratory support Index (RI) at baseline was tested using linear regression models. RI  
694 groups were as follows: RI 0: no respiratory support required, RI 1:  $\leq 5$  Litres of oxygen on nasal  
695 cannula or mask, RI 2:  $> 5$  Litres of oxygen on nasal cannula or mask, RI 3: non-invasive  
696 ventilation/high-flow nasal cannula, or RI 4: intubation. To handle multiple time point  
697 measurements, association of CFHR5 plasma levels with RI was further investigated using all  
698 available longitudinal measurements using a linear mixed effect model as implemented in the  
699 *nlme* R package. Analyses were adjusted for age, sex and body mass index.

#### 700 **CFHR5 mRNA expression across human organs**

701 As part of the Human Protein Atlas (HPA, [www.proteinatlas.org/](http://www.proteinatlas.org/)), the average TPM value of all  
702 individual samples for each human tissue in both the HPA and Genotype-Tissue Expression  
703 Project (GTEx) transcriptomics datasets were used to estimate the respective gene expression  
704 levels. To be able to combine the datasets into consensus transcript expression levels, a pipeline  
705 was set up to normalize the data for all samples. In brief, all TPM values per sample were scaled  
706 to a sum of 1 million TPM (denoted pTPM) to compensate for the non-coding transcripts that had  
707 been previously removed. Next, all TPM values of all samples within each data source were  
708 normalized separately using Trimmed mean of M values to allow for between-sample  
709 comparisons. The resulting normalized transcript expression values (nTPM) were calculated for  
710 each gene in each sample. For further details see [www.proteinatlas.org/about/assays+annotation](http://www.proteinatlas.org/about/assays+annotation-normalization_rna)  
711 [- normalization\\_rna](http://www.proteinatlas.org/about/assays+annotation-normalization_rna). Analysis of liver single cell transcriptomes and visualisation was performed  
712 as part of the HPA single-cell transcriptomics map [41] from data generated in [84].

#### 713 **CFHR5 mRNA co-expression analysis**

714 Liver bulk RNAseq data analysed in this study was part of the Genotype-Tissue Expression  
715 (GTEx) Project ([gtexportal.org](http://gtexportal.org)) [85] (dbGaP Accession phs000424.v7.p2) (n=226). Pearson

716 correlation coefficients were calculated between *CFHR5* expression values and those for all other  
717 mapped protein coding genes across the sample set. Weighted correlation network (WGCNA)  
718 analysis: The R package WGCNA was used to perform co-expression network analysis for gene  
719 clustering, on log<sub>2</sub> expression values. The analysis was performed according to  
720 recommendations in the WGCNA manual. Genes with too many missing values were excluded  
721 using the goodSamplesGenes() function. The remaining genes were used to cluster the samples,  
722 and obvious outlier samples were excluded. Using these genes and samples a soft-thresholding  
723 power was selected and the networks were constructed using a minimum module size of 15 and  
724 merging threshold of 0.05. Eigengenes were calculated from the resulting clusters and eigengene  
725 dendrograms were constructed using the plotEigengeneNetworks() function.

#### 726 **Measurement of Thrombin Generation Potential**

727 Thrombin generation potential (TGP) was measured in fresh frozen platelet poor plasma (PPP)  
728 using the Calibrated Automated Thrombogram (CAT®) method according to the manufacturer's  
729 instructions. Analyses in MARTHA and RETROVE are as previously described [80, 86].

730 Output parameters recorded were lagtime (min), the time to the initial generation of thrombin after  
731 induction; Endogenous Thrombin Potential (ETP)(nmol/min), equal to the area under the  
732 Thrombogram curve; Peak (nmol/L), the maximum amount of thrombin produced after induction  
733 by 5pM tissue factor.

#### 734 **Effect of recombinant CFHR5 on platelet activation**

735 Blood was drawn from healthy volunteers free from any anti-platelet therapy for at least 10 days  
736 and anticoagulated with sodium citrate. All donors signed informed consent, in accordance with  
737 approval of the Human Ethics Committee of the Medical University of Vienna (EK237/2004) and  
738 the Declaration of Helsinki. Whole blood was centrifuged (120 g, 20 minutes, room temperature)  
739 and platelet-rich plasma (PRP) harvested. To obtain isolated platelets, PRP was diluted with PBS  
740 and treated with PGI<sub>2</sub> (100 ng/ml), centrifuged for 90 sec at 3000 x g and platelets were  
741 resuspended in PBS. This step was repeated twice. Platelet-rich plasma (PRP) or isolated

742 platelets were incubated with recombinant CFHR5 in PBS (6 µg/ml, 3845-F5, R&D systems) or  
743 PBS alone for 10 minutes before treatment with varying concentrations of ADP (1-5 µM), TRAP-  
744 6 (3-15 µM) or convluxin (1-6 ng/ml) for 15 minutes. Platelets were subsequently incubated with  
745 primary antibodies: anti-human CD62P-AF647 (AK4), anti-human CD63-PE (H5C6) or anti-  
746 human CD41/CD61-FITC (PAC-1) (all Biolegend) for 20 minutes, washed (PBS then 500 g for  
747 10 minutes), then fixed with 1 % paraformaldehyde and incubated with Alexa Fluor 647-  
748 streptavidin (Jackson Immuno Research, Ely, UK) for 20 minutes. Samples were analysed by flow  
749 cytometry (Cytotflex, Beckman Coulter GmbH, Krefeld, Germany) and data processed using  
750 Cytexpert (Beckman Coulter GmbH, Krefeld, Germany).

### 751 **Genome wide genotyping methods**

752 *RETROVE* samples were typed with the Illumina Infinium Global Screening Array (GSA) v2.0  
753 array at the Spanish National Cancer Research Centre in the Human Genotyping lab, a member  
754 of CeGen. After genotyping, all monomorphic and unannotated variants were removed as well as  
755 polymorphisms with call rate <95% and those whose genotype distributions deviate from Hardy-  
756 Weinberg equilibrium at  $p < 0.000001$ . Remaining polymorphisms were then imputed using the  
757 TOPMed r2 reference panel using Eagle v2.4. *FARIVE* participants were genotyped using the  
758 Illumina Infinium Global Screening Array v3.0 (GSAv3.0) microarray at the Centre National de  
759 Recherche en Génomique Humaine (CNRGH). A control quality has been performed on  
760 individuals and genetic variants using Plink v1.9 and the R software v3.6.2 (Chang et al., 2015).  
761 Individuals with at least one of the following criteria were excluded: discordant sex information  
762 (N=20), relatedness individuals (N=9) identified by pairwise clustering of identity by state  
763 distances (IBS), genotyping call rate lower than 99% (N=5), heterozygosity rate higher/lower than  
764 the average rate +/- 3 standard deviation (N=34). These criteria led to a final sample composed  
765 of 1,266 individuals. Among the 730,059 genotyped variants, we excluded 145,238 variants with  
766 incorrect annotation, 656 variants with deviation from Hardy-Weinberg equilibrium (HWE) in  
767 controls using the statistical threshold of  $p < 10^{-6}$ , 47,286 variants with a Minor Allele Count (MAC)

768 lower than 20, 1,774 variants with a call rate lower than 95%. Finally, 535,105 markers passed  
769 the control quality and were used for the imputation. The Imputation was performed with Minimac4  
770 using the 1000 Genomes phase 3 version 5 reference panel [87]. *MARTHA* samples were typed  
771 with Illumina Human 610-/ 660W- Quad Beadchip. Quality control procedures of produced  
772 genome-wide genotype data in *MARTHA* participants have already been extensively described  
773 [88-90]. After further exclusion of individuals with cancer and systemic lupus erythematosus, 1525  
774 participants remained for association testing.

#### 775 **Genome Wide Association Study on CFHR5 plasma levels**

776 All SNPs with imputation quality criterion greater than 0.30 and minor allele frequency greater  
777 than 0.01 in each participating cohorts were tested for association with CFHR5 plasma levels.  
778 Associations were assessed using a linear regression model adjusted for age, sex and study-  
779 specific principal components derived from genome-wide genotype data. Results obtained in the  
780 different contributing cohorts were then meta-analyzed through a random effect model as  
781 implemented in the GWAMA software [91].



782 **ACKNOWLEDGEMENTS**

783 The study was supported by grants from Stockholm City Council (SLL) to JO (2017-0842/0587,  
784 2020-0346), from Familjen Erling Personssons Stiftelse to MU, Knut och Alice Wallenberg  
785 foundation to JO (2020.0182, 2020.0241), from Swedish Heart Lung Foundation to LB (20170759,  
786 20170537), National Research Council (VR) to LB (2019-01493), HelseNord to JO (HNF1544-  
787 20). The Human Protein Atlas (HPA) is funded by The Knut and Alice Wallenberg Foundation.

788 *MARTHA* and *FARIVE* related genetics research programs were funded by the  
789 GENMED Laboratory of Excellence on Medical Genomics [ANR-10-LABX-0013], a research  
790 program managed by the National Research Agency (ANR) as part of the French Investment for  
791 the Future, and supported by the French INvestigation Network on Venous Thrombo-Embolism  
792 (*INNOVTE*). *MARTHA* and *FARIVE* genetic data analyses benefit from the technical support of  
793 the CBiB computing centre of the University of Bordeaux. GM and D-AT are supported by the  
794 EPIDEMIOIOM-VT Senior Chair from the University of Bordeaux initiative of excellence IdEX. GM  
795 benefited from the EUR DPH, a PhD program supported within the framework of the PIA3  
796 (Investment for the future). Project reference 17-EURE-0019. The *RETROVE* study was  
797 supported by grants PI12/00612 and PI15/0026. Genotyping of the *RETROVE* samples was  
798 supported by grant PT17/0019, of the PE I+D+i 2013-2016, funded by ISCIII and ERDF.

799 **Data usage:** We used data from the Genotype-Tissue Expression (GTEx) Project ([gtexportal.org](http://gtexportal.org))  
800 [40]. The GTEx project was supported by the Office of the Director of the National Institutes of  
801 Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

802 **Figures:** Some parts of the figures were created with BioRender.com.

803 **References**

- 804 1. Johansson, M., L. Johansson, and M. Lind, *Incidence of venous thromboembolism in*  
805 *northern Sweden (VEINS): a population-based study*. *Thromb J*, 2014. **12**(1): p. 6.
- 806 2. Heit, J.A., *Epidemiology of venous thromboembolism*. *Nature reviews. Cardiology*, 2015.  
807 **12**(8): p. 464-74.
- 808 3. Sogaard, K.K., et al., *30-year mortality after venous thromboembolism: a population-*  
809 *based cohort study*. *Circulation*, 2014. **130**(10): p. 829-36.
- 810 4. Martinez, C., et al., *Epidemiology of first and recurrent venous thromboembolism: A*  
811 *population-based cohort study in patients without active cancer*. *Thrombosis and*  
812 *Haemostasis*, 2014. **112**(2): p. 255-263.
- 813 5. Wells, P.S., et al., *Evaluation of D-dimer in the diagnosis of suspected deep-vein*  
814 *thrombosis*. *N Engl J Med*, 2003. **349**(13): p. 1227-35.
- 815 6. Wells, P.S., et al., *Excluding pulmonary embolism at the bedside without diagnostic*  
816 *imaging: management of patients with suspected pulmonary embolism presenting to the*  
817 *emergency department by using a simple clinical model and d-dimer*. *Ann Intern Med*,  
818 2001. **135**(2): p. 98-107.
- 819 7. Mittadodla, P.S., et al., *CT pulmonary angiography: an over-utilized imaging modality in*  
820 *hospitalized patients with suspected pulmonary embolism*. *J Community Hosp Intern*  
821 *Med Perspect*, 2013. **3**(1).
- 822 8. Sikkens, J.J., et al., *How Much Overtesting Is Needed to Safely Exclude a Diagnosis? A*  
823 *Different Perspective on Triage Testing Using Bayes' Theorem*. *PLoS One*, 2016. **11**(3):  
824 p. e0150891.
- 825 9. Pernod, G., et al., *D-Dimer Use and Pulmonary Embolism Diagnosis in Emergency*  
826 *Units: Why Is There Such a Difference in Pulmonary Embolism Prevalence between the*  
827 *United States of America and Countries Outside USA?* *PLoS One*, 2017. **12**(1): p.  
828 e0169268.
- 829 10. Francis, S., et al., *Highly Elevated Quantitative D-Dimer Assay Values Increase the*  
830 *Likelihood of Venous Thromboembolism*. *TH Open*, 2019. **3**(1): p. e2-e9.
- 831 11. Jacobs, B., A. Obi, and T. Wakefield, *Diagnostic biomarkers in venous thromboembolic*  
832 *disease*. *J Vasc Surg Venous Lymphat Disord*, 2016. **4**(4): p. 508-17.
- 833 12. Eichinger, S., et al., *Risk assessment of recurrence in patients with unprovoked deep vein*  
834 *thrombosis or pulmonary embolism: the Vienna prediction model*. *Circulation*, 2010.  
835 **121**(14): p. 1630-6.
- 836 13. Tosetto, A., et al., *Predicting disease recurrence in patients with previous unprovoked*  
837 *venous thromboembolism: a proposed prediction score (DASH)*. *J Thromb Haemost*,  
838 2012. **10**(6): p. 1019-25.
- 839 14. Verhovsek, M., et al., *Systematic review: D-dimer to predict recurrent disease after*  
840 *stopping anticoagulant therapy for unprovoked venous thromboembolism*. *Ann Intern*  
841 *Med*, 2008. **149**(7): p. 481-90, W94.
- 842 15. Rodger, M.A., et al., *Identifying unprovoked thromboembolism patients at low risk for*  
843 *recurrence who can discontinue anticoagulant therapy*. *Canadian Medical Association*  
844 *Journal*, 2008. **179**(5): p. 417-26.
- 845 16. Baglin, T., et al., *Unprovoked recurrent venous thrombosis: prediction by D-dimer and*  
846 *clinical risk factors*. *J Thromb Haemost*, 2008. **6**(4): p. 577-82.

- 847 17. Bruinstroop, E., et al., *Elevated D-dimer levels predict recurrence in patients with*  
848 *idiopathic venous thromboembolism: a meta-analysis*. J Thromb Haemost, 2009. **7**(4): p.  
849 611-8.
- 850 18. Douketis, J., et al., *Patient-level meta-analysis: effect of measurement timing, threshold,*  
851 *and patient age on ability of D-dimer testing to assess recurrence risk after unprovoked*  
852 *venous thromboembolism*. Ann Intern Med, 2010. **153**(8): p. 523-31.
- 853 19. Lindstrom, S., et al., *Genomic and transcriptomic association studies identify 16 novel*  
854 *susceptibility loci for venous thromboembolism*. Blood, 2019. **134**(19): p. 1645-1657.
- 855 20. Klarin, D., et al., *Genome-wide association analysis of venous thromboembolism*  
856 *identifies new risk loci and genetic overlap with arterial vascular disease*. Nat Genet,  
857 2019. **51**(11): p. 1574-1579.
- 858 21. Thibord, F.e.a., <https://doi.org/10.1101/2022.03.04.22271003>.
- 859 22. Morange, P.E. and D.A. Tregouet, *Current knowledge on the genetics of incident venous*  
860 *thrombosis*. J Thromb Haemost, 2013. **11 Suppl 1**: p. 111-21.
- 861 23. Martinelli, I., V. De Stefano, and P.M. Mannucci, *Inherited risk factors for venous*  
862 *thromboembolism*. Nat Rev Cardiol, 2014. **11**(3): p. 140-56.
- 863 24. Goldhaber, S.Z., *Risk factors for venous thromboembolism*. J Am Coll Cardiol, 2010.  
864 **56**(1): p. 1-7.
- 865 25. Crous-Bou, M., L.B. Harrington, and C. Kabrhel, *Environmental and Genetic Risk*  
866 *Factors Associated with Venous Thromboembolism*. Semin Thromb Hemost, 2016. **42**(8):  
867 p. 808-820.
- 868 26. Jensen, S.B., et al., *Discovery of novel plasma biomarkers for future incident venous*  
869 *thromboembolism by untargeted synchronous precursor selection mass spectrometry*  
870 *proteomics*. Journal of Thrombosis and Haemostasis, 2018. **16**: p. 1763-1774.
- 871 27. Bruzelius, M., et al., *PDGFB, a new candidate plasma biomarker for venous*  
872 *thromboembolism: results from the VEREMA affinity proteomics study*. Blood, 2016.  
873 **128**(23): p. 59-67.
- 874 28. Blostein, M.D., et al., *Elevated plasma gas6 levels are associated with venous*  
875 *thromboembolic disease*. J Thromb Thrombolysis, 2011. **32**(3): p. 272-8.
- 876 29. Song, Y., et al., *Increased expressions of integrin subunit beta1, beta2 and beta3 in*  
877 *patients with venous thromboembolism: new markers for venous thromboembolism*. Int J  
878 Clin Exp Med, 2014. **7**(9): p. 2578-84.
- 879 30. Memon, A.A., et al., *Identification of novel diagnostic biomarkers for deep venous*  
880 *thrombosis*. British Journal of Haematology, 2018. **181**: p. 378-385.
- 881 31. Deutsch, E.W., et al., *Advances and Utility of the Human Plasma Proteome*. J Proteome  
882 Res, 2021. **20**(12): p. 5241-5263.
- 883 32. Butler, L.M., et al., *Analysis of Body-wide Unfractionated Tissue Data to Identify a Core*  
884 *Human Endothelial Transcriptome*. Cell Syst, 2016. **3**(3): p. 287-301 e3.
- 885 33. Mathews, J.A., et al., *Considerations for Soluble Protein Biomarker Blood Sample*  
886 *Matrix Selection*. AAPS J, 2020. **22**(2): p. 38.
- 887 34. Baker, M., *Reproducibility crisis: Blame it on the antibodies*. Nature, 2015. **521**(7552): p.  
888 274-6.
- 889 35. Weller, M.G., *Quality Issues of Research Antibodies*. Anal Chem Insights, 2016. **11**: p.  
890 21-7.
- 891 36. Fredolini, C., et al., *Systematic assessment of antibody selectivity in plasma based on a*  
892 *resource of enrichment profiles*. Sci Rep, 2019. **9**(1): p. 8324.

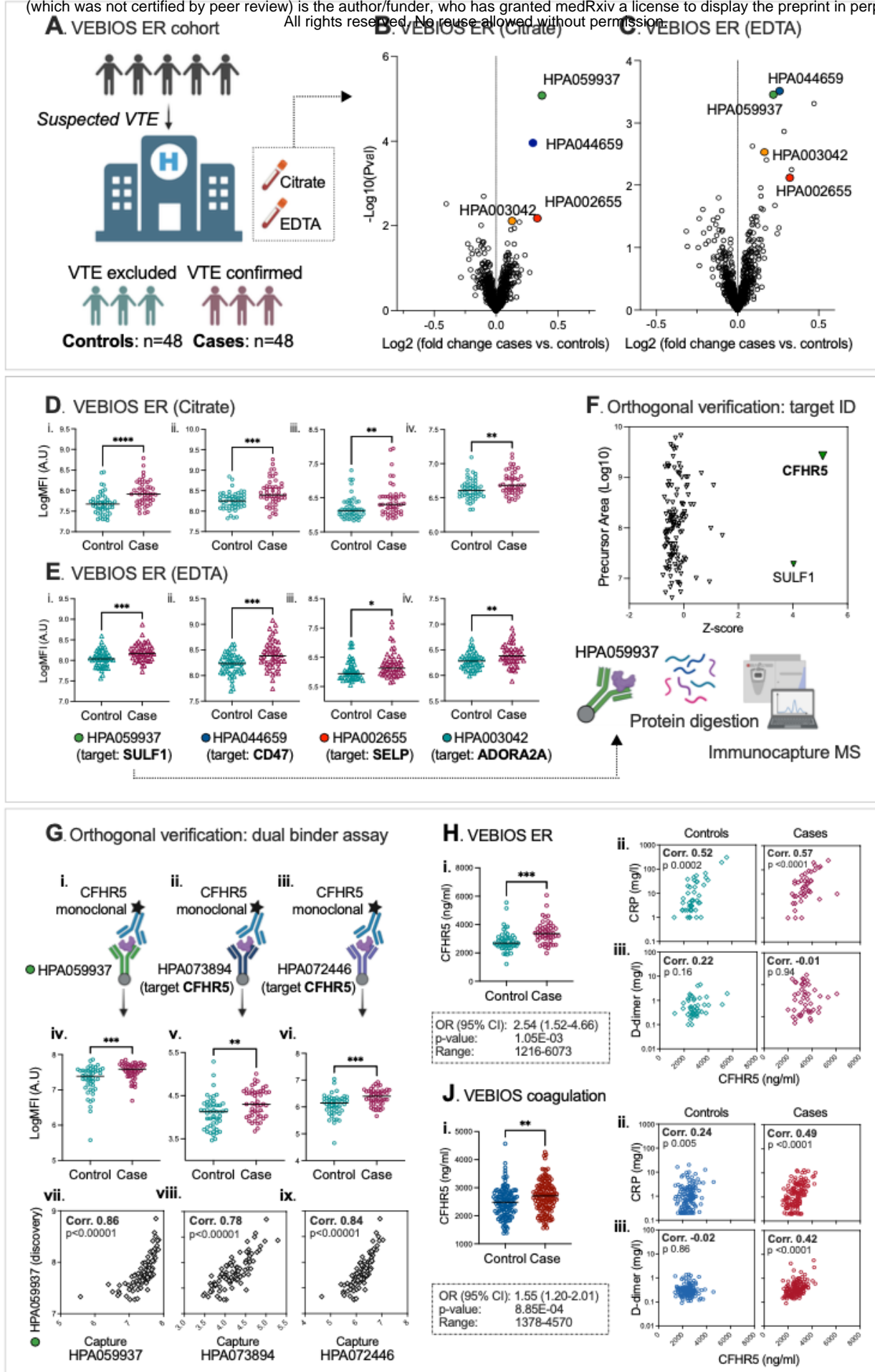
- 893 37. Schwenk, J.M., et al., *The Human Plasma Proteome Draft of 2017: Building on the*  
894 *Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays.* J  
895 Proteome Res, 2017. **16**(12): p. 4299-4310.
- 896 38. Desiere, F., et al., *The PeptideAtlas project.* Nucleic Acids Res, 2006. **34**(Database  
897 issue): p. D655-8.
- 898 39. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015.  
899 **347**(6220): p. 1260419.
- 900 40. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013.  
901 **45**(6): p. 580-5.
- 902 41. Karlsson, M., et al., *A single-cell type transcriptomics map of human tissues.* Sci Adv,  
903 2021. **7**(31).
- 904 42. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene*  
905 *Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
- 906 43. Gene Ontology, C., *The Gene Ontology resource: enriching a Gold mine.* Nucleic Acids  
907 Res, 2021. **49**(D1): p. D325-D334.
- 908 44. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network*  
909 *analysis.* BMC Bioinformatics, 2008. **9**: p. 559.
- 910 45. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased*  
911 *coverage, supporting functional discovery in genome-wide experimental datasets.*  
912 Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
- 913 46. Norgaard, I., S.F. Nielsen, and B.G. Nordestgaard, *Complement C3 and High Risk of*  
914 *Venous Thromboembolism: 80517 Individuals from the Copenhagen General Population*  
915 *Study.* Clin Chem, 2016. **62**(3): p. 525-34.
- 916 47. Farm, M., et al., *Age-adjusted D-dimer cut-off leads to more efficient diagnosis of venous*  
917 *thromboembolism in the emergency department: a comparison of four assays.* J Thromb  
918 Haemost, 2018. **16**(5): p. 866-875.
- 919 48. Zhu, T., et al., *Association of influenza vaccination with reduced risk of venous*  
920 *thromboembolism.* Thromb Haemost, 2009. **102**(6): p. 1259-64.
- 921 49. Llobet, D., et al., *Platelet hyperaggregability and venous thrombosis risk: results from*  
922 *the RETROVE project.* Blood Coagul Fibrinolysis, 2021. **32**(2): p. 122-131.
- 923 50. Oudot-Mellakh, T., et al., *Genome wide association study for plasma levels of natural*  
924 *anticoagulant inhibitors and protein C anticoagulant pathway: the MARTHA project.* Br  
925 J Haematol, 2012. **157**(2): p. 230-9.
- 926 51. Sauter, R.J., et al., *Functional Relevance of the Anaphylatoxin Receptor C3aR for*  
927 *Platelet Function and Arterial Thrombus Formation Marks an Intersection Point*  
928 *Between Innate Immunity and Thrombosis.* Circulation, 2018. **138**(16): p. 1720-1735.
- 929 52. Afshar-Kharghan, V., *Complement and clot.* Blood, 2017. **129**(16): p. 2214-2215.
- 930 53. Subramaniam, S., et al., *Distinct contributions of complement factors to platelet*  
931 *activation and fibrin formation in venous thrombus development.* Blood, 2017. **129**(16):  
932 p. 2291-2302.
- 933 54. McRae, J.L., et al., *Human factor H-related protein 5 has cofactor activity, inhibits C3*  
934 *convertase activity, binds heparin and C-reactive protein, and associates with*  
935 *lipoprotein.* J Immunol, 2005. **174**(10): p. 6250-6.
- 936 55. Wang, H., et al., *D-dimer, thrombin generation, and risk of a first venous thrombosis in*  
937 *the elderly.* Res Pract Thromb Haemost, 2021. **5**(5): p. e12536.

- 938 56. Magro, C., et al., *Complement associated microvascular injury and thrombosis in the*  
939 *pathogenesis of severe COVID-19 infection: A report of five cases.* Transl Res, 2020.  
940 **220**: p. 1-13.
- 941 57. Wu, Z. and J.M. McGoogan, *Characteristics of and Important Lessons From the*  
942 *Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of*  
943 *72314 Cases From the Chinese Center for Disease Control and Prevention.* JAMA,  
944 2020. **323**(13): p. 1239-1242.
- 945 58. Poissy, J., et al., *Pulmonary Embolism in Patients With COVID-19: Awareness of an*  
946 *Increased Prevalence.* Circulation, 2020. **142**(2): p. 184-186.
- 947 59. von Meijenfeldt, F.A., et al., *Prothrombotic changes in patients with COVID-19 are*  
948 *associated with disease severity and mortality.* Res Pract Thromb Haemost, 2021. **5**(1): p.  
949 132-141.
- 950 60. Audu, C.O., et al., *Inflammatory biomarkers in deep venous thrombosis organization,*  
951 *resolution, and post-thrombotic syndrome.* J Vasc Surg Venous Lymphat Disord, 2020.  
952 **8**(2): p. 299-305.
- 953 61. Mosevoll, K.A., et al., *Altered plasma levels of cytokines, soluble adhesion molecules*  
954 *and matrix metalloproteases in venous thrombosis.* Thromb Res, 2015. **136**(1): p. 30-9.
- 955 62. van Hylckama Vlieg, A., et al., *The risk of a first and a recurrent venous thrombosis*  
956 *associated with an elevated D-dimer level and an elevated thrombin potential: results of*  
957 *the THE-VTE study.* J Thromb Haemost, 2015. **13**(9): p. 1642-52.
- 958 63. McRae, J.L., et al., *Human factor H-related protein 5 (FHR-5). A new complement-*  
959 *associated protein.* J Biol Chem, 2001. **276**(9): p. 6747-54.
- 960 64. Cserhalmi, M., et al., *Regulation of regulators: Role of the complement factor H-related*  
961 *proteins.* Semin Immunol, 2019. **45**: p. 101341.
- 962 65. Zipfel, P.F., et al., *Factor H and disease: a complement regulator affects vital body*  
963 *functions.* Mol Immunol, 1999. **36**(4-5): p. 241-8.
- 964 66. Murphy, B., et al., *Factor H-related protein-5: a novel component of human glomerular*  
965 *immune deposits.* Am J Kidney Dis, 2002. **39**(1): p. 24-7.
- 966 67. Csincsi, A.I., et al., *Factor H-related protein 5 interacts with pentraxin 3 and the*  
967 *extracellular matrix and modulates complement activation.* J Immunol, 2015. **194**(10): p.  
968 4963-73.
- 969 68. Markiewski, M.M., et al., *Complement and coagulation: strangers or partners in crime?*  
970 Trends Immunol, 2007. **28**(4): p. 184-92.
- 971 69. Wu, Z., et al., *Time Course of a Second Outbreak of COVID-19 in Beijing, China, June-*  
972 *July 2020.* JAMA, 2020. **324**(14): p. 1458-1459.
- 973 70. Marini, J.J. and L. Gattinoni, *Management of COVID-19 Respiratory Distress.* JAMA,  
974 2020. **323**(22): p. 2329-2330.
- 975 71. Varga, Z., et al., *Endothelial cell infection and endotheliitis in COVID-19.* Lancet, 2020.  
976 **395**(10234): p. 1417-1418.
- 977 72. Ma, L., et al., *Increased complement activation is a distinctive feature of severe SARS-*  
978 *CoV-2 infection.* bioRxiv, 2021.
- 979 73. Klok, F.A., et al., *Incidence of thrombotic complications in critically ill ICU patients with*  
980 *COVID-19.* Thromb Res, 2020. **191**: p. 145-147.
- 981 74. Joshi, A. and M. Mayr, *In Aptamers They Trust: The Caveats of the SOMAscan*  
982 *Biomarker Discovery Platform from SomaLogic.* Circulation, 2018. **138**(22): p. 2482-  
983 2485.

- 984 75. Pietzner, M., et al., *Synergistic insights into human health from aptamer- and antibody-*  
985 *based proteomic profiling*. Nat Commun, 2021. **12**(1): p. 6822.
- 986 76. Olson, N.C., et al., *Soluble Urokinase Plasminogen Activator Receptor: Genetic*  
987 *Variation and Cardiovascular Disease Risk in Black Adults*. Circ Genom Precis Med,  
988 2021. **14**(6): p. e003421.
- 989 77. Zhu, T., et al., *Association of influenza vaccination with reduced risk of venous*  
990 *thromboembolism*. Thrombosis and Haemostasis, 2009. **102**(12): p. 1259-1264.
- 991 78. Trégouët, D.A., et al., *Common susceptibility alleles are unlikely to contribute as*  
992 *strongly as the FV and ABO loci to VTE risk: Results from a GWAS approach*. Blood,  
993 2009. **113**(21): p. 5298-5303.
- 994 79. Vazquez-Santiago, M., et al., *Platelet count and plateletcrit are associated with an*  
995 *increased risk of venous thrombosis in females. Results from the RETROVE study*.  
996 Thromb Res, 2017. **157**: p. 162-164.
- 997 80. Rocanin-Arjo, A., et al., *A meta-analysis of genome-wide association studies identifies*  
998 *ORM1 as a novel gene controlling thrombin generation potential*. Blood, 2014. **123**(5): p.  
999 777-785.
- 1000 81. Dieterle, F., et al., *Probabilistic quotient normalization as robust method to account for*  
1001 *dilution of complex biological mixtures. Application in 1H NMR metabolomics*. Anal  
1002 Chem, 2006. **78**(13): p. 4281-90.
- 1003 82. Hong, M.G., et al., *Multidimensional Normalization to Minimize Plate Effects of*  
1004 *Suspension Bead Array Data*. J Proteome Res, 2016. **15**(10): p. 3473-3480.
- 1005 83. Ihaka, R.G., R, R: *A Language for Data Analysis and Graphics*. Journal of  
1006 Computational and Graphical Statistics, 1996. **5**(3): p. 299-314.
- 1007 84. MacParland, S.A., et al., *Single cell RNA sequencing of human liver reveals distinct*  
1008 *intrahepatic macrophage populations*. Nat Commun, 2018. **9**(1): p. 4383.
- 1009 85. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot*  
1010 *analysis: multitissue gene regulation in humans*. Science, 2015. **348**(6235): p. 648-60.
- 1011 86. Martin-Fernandez, L., et al., *Genetic Determinants of Thrombin Generation and Their*  
1012 *Relation to Venous Thrombosis: Results from the GAIT-2 Project*. PLoS One, 2016.  
1013 **11**(1): p. e0146922.
- 1014 87. Das, S., et al., *Next-generation genotype imputation service and methods*. Nat Genet,  
1015 2016. **48**(10): p. 1284-1287.
- 1016 88. Germain, M., et al., *Genetics of venous thrombosis: insights from a new genome wide*  
1017 *association study*. PLoS One, 2011. **6**(9): p. e25581.
- 1018 89. Antoni, G., et al., *Combined analysis of three genome-wide association studies on vWF*  
1019 *and FVIII plasma levels*. BMC Med Genet, 2011. **12**: p. 102.
- 1020 90. Germain, M., et al., *Meta-analysis of 65,734 individuals identifies TSPAN15 and*  
1021 *SLC44A2 as two susceptibility loci for venous thromboembolism*. Am J Hum Genet,  
1022 2015. **96**(4): p. 532-42.
- 1023 91. Magi, R. and A.P. Morris, *GWAMA: software for genome-wide association meta-*  
1024 *analysis*. BMC Bioinformatics, 2010. **11**: p. 288.
- 1025

Figure 1

medRxiv preprint doi: <https://doi.org/10.1101/2022.04.20.22274046>; this version posted April 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



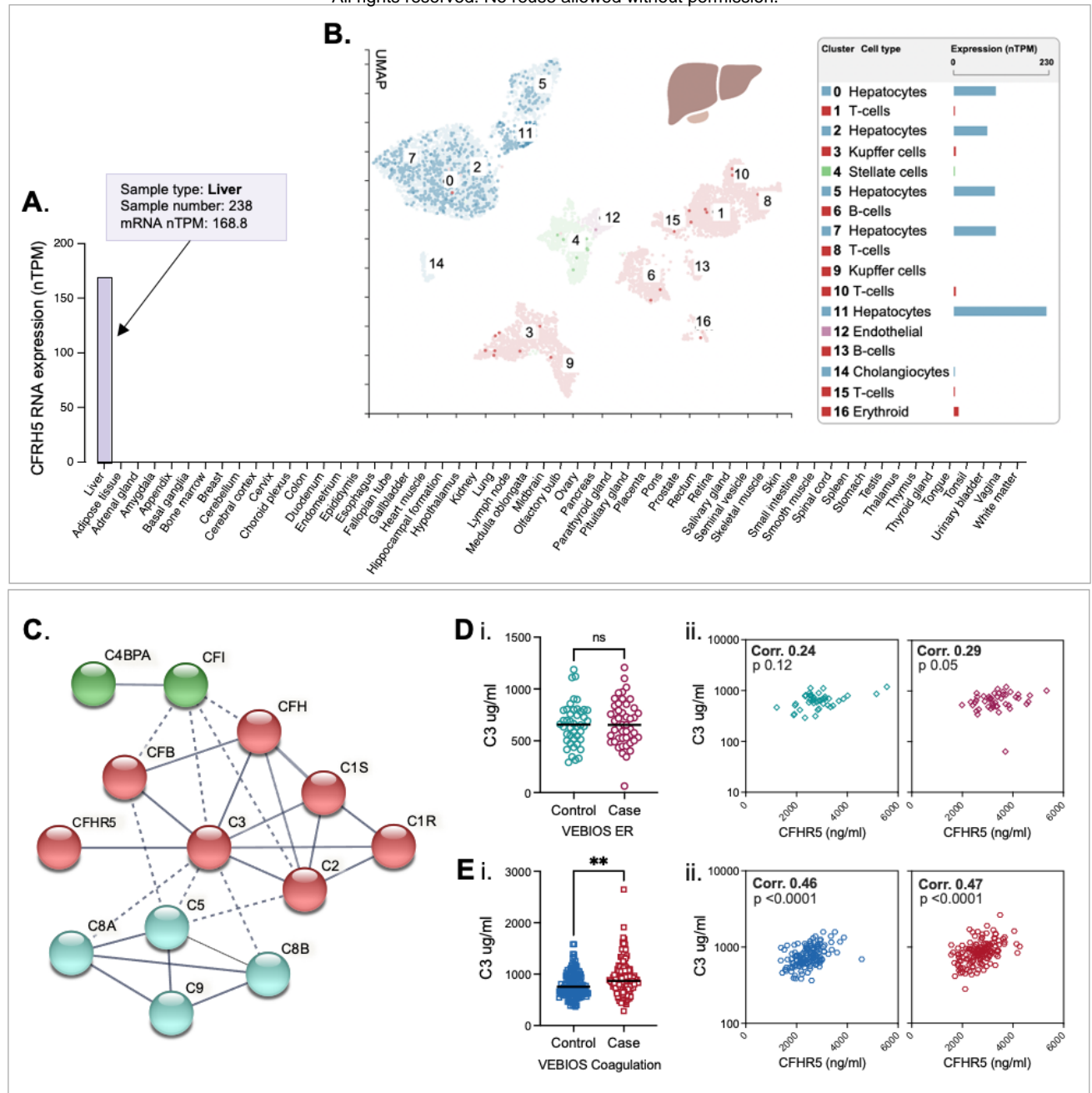
## Figure 1. Plasma proteomics profiling identifies CFHR5 as associated with VTE

(A) Overview of the VEBIOS ER discovery cohort. 758 HPA antibodies, targeting 408 candidate proteins, were used to analyse plasma samples using affinity proteomics. Log fold changes in antibody MFI (mean fluorescent intensity) signal were calculated between VTE cases and controls in (B) citrate or (C) EDTA anticoagulated plasma; coloured circles indicate antibodies that generated signals significantly associated with VTE in both. MFI signals generated by these antibodies for controls and cases in (D) citrate plasma and (E) EDTA plasma. (F) Immunocapture-mass spectrometry identification of protein targets of HPA059937. (G) Dual binder assays were developed using an anti-CFHR5 detection antibody, combined with (i) HPA059937 (raised against SULF1) (ii), anti-CFHR5 HPA073894 or (iii) anti-CFHR5 HPA072446 as capture antibodies. CFHR5 levels in the citrated plasma samples were re-analysed, using the respective dual binder assays, to determine (vii-ix) levels (MFI) in controls vs. cases and (vii-ix) the correlation between the signal and those generated by the original single binder assay using HPA059937. Dual binder assay using capture antibody HPA072446 with a recombinant protein standard was used for absolute quantification of CFHR5 in samples from: (H) *VEBIOS ER* and (J) *VEBIOS Coagulation*. CFHR5 concentration was (i) measured in controls and cases, with associated OR (odds ratio per 1 standard deviation increase) or (ii) used to determine the correlation with C-reactive protein (CRP), or (iii) d-dimer concentration. \*\*\* $p < 0.0001$ , \*\* $p < 0.001$ , \* $p < 0.01$ .



Figure 2

medRxiv preprint doi: <https://doi.org/10.1101/2022.04.20.22274046>; this version posted April 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



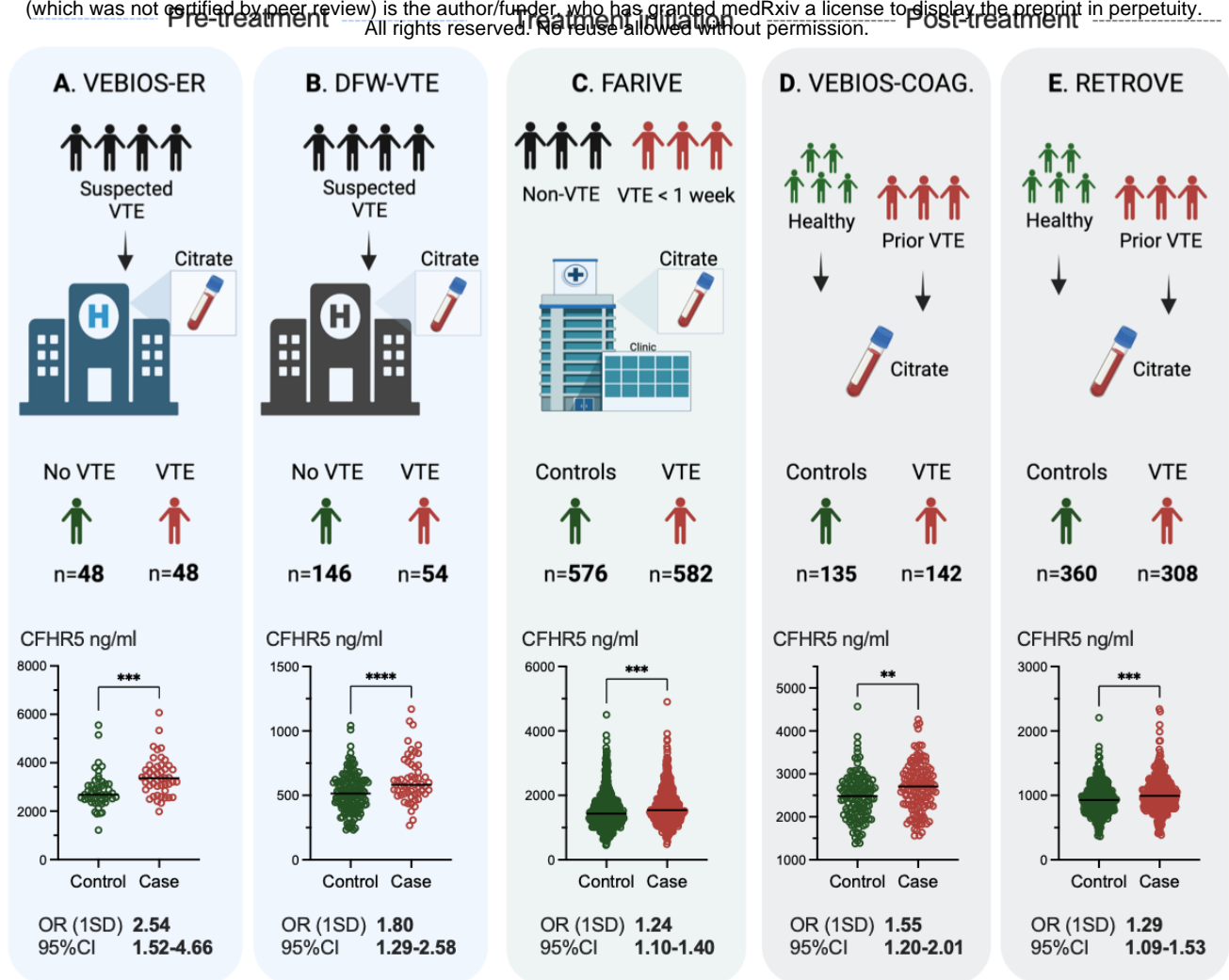
**Figure 2: *CFHR5* is expressed in hepatocytes and is VTE-associated independent of C3**

(A) mRNA expression of *CFHR5* across 55 different human tissue types. (B) Expression of *CFHR5* in different liver cell types, analysed by ssRNAseq. (C) STRING protein-protein interaction analysis for genes identified as potentially co-expressed with *CFHR5* in liver by correlation-based analysis of bulk mRNAseq. Coloured circles indicate closest network clusters. Complement component 3 (C3) concentration was measured in plasma from (D) *VEBIOS ER* or (E) *VEBIOS coagulation* to determine: (i) differences between controls and cases, or (ii) correlation with *CFHR5* in controls (left) or cases (right). \*\*p<0.01.

CFH; complement factor H, C1S; complement component 1, C1R; complement component 1, C2; complement component 2, C8a; complement component C8 alpha chain, C8b; complement component C8 beta chain, C5 complement component 5, C9; complement component 9

Figure 3

medRxiv preprint doi: <https://doi.org/10.1101/2022.04.20.22274046>; this version posted April 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

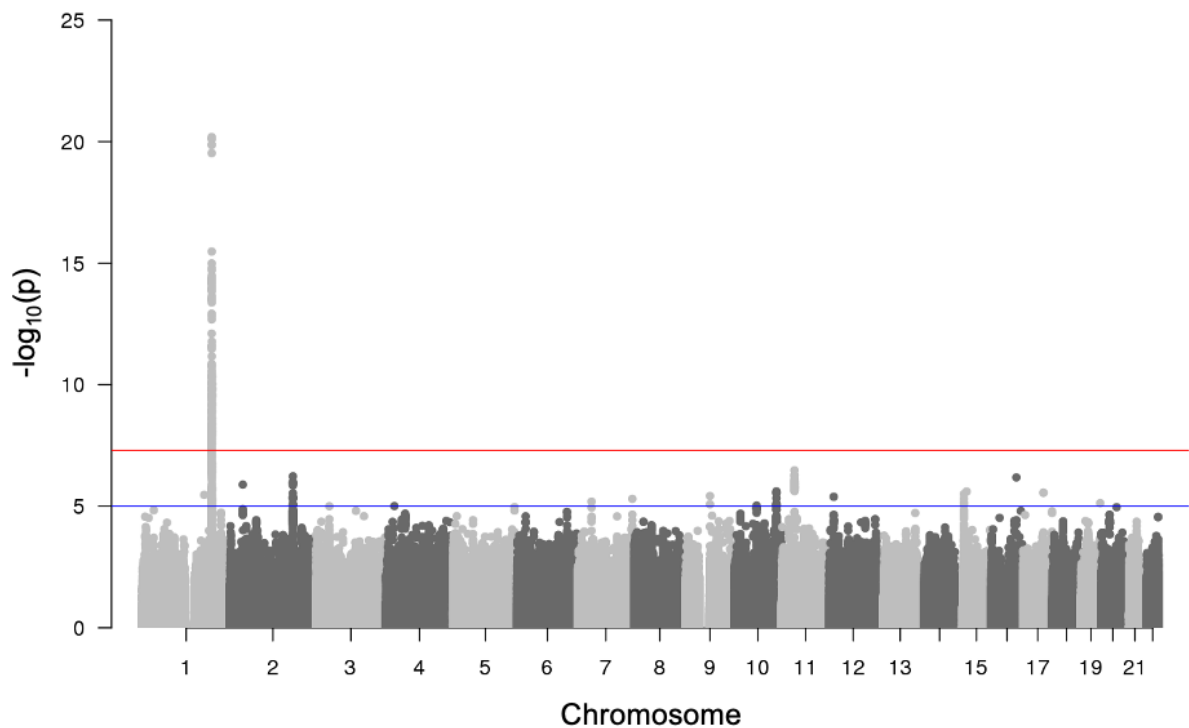


**Figure 3: CFHR5 concentration is associated with VTE in 5 independent studies.** Plasma samples were generated as part of: (A) the Swedish *VEBIOS ER* or (B) the Swedish *DFW-VTE* study, both of which recruited patients presenting with suspected VTE. Samples were drawn pre-treatment, and cases and controls were identified based on confirmed or ruled out diagnosis. (C) The French *FARIVE* study recruited patients with confirmed acute VTE, with controls recruited from hospital patients treated for non-VTE causes. Samples were drawn within one week from diagnosis, during initiation of treatment. (D) The Swedish *VEBIOS Coagulation* or (E), Spanish *RETROVE* study recruited cases from patients who had a prior first time VTE, sampled post-treatment (6-12 months anticoagulants), with healthy controls recruited from the general population. CFHR5 concentration was measured in the respective samples using a dual binder assay \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.001$  OR (1SD) = Odds ratio for 1 standard deviation elevation. CI = confidence interval.

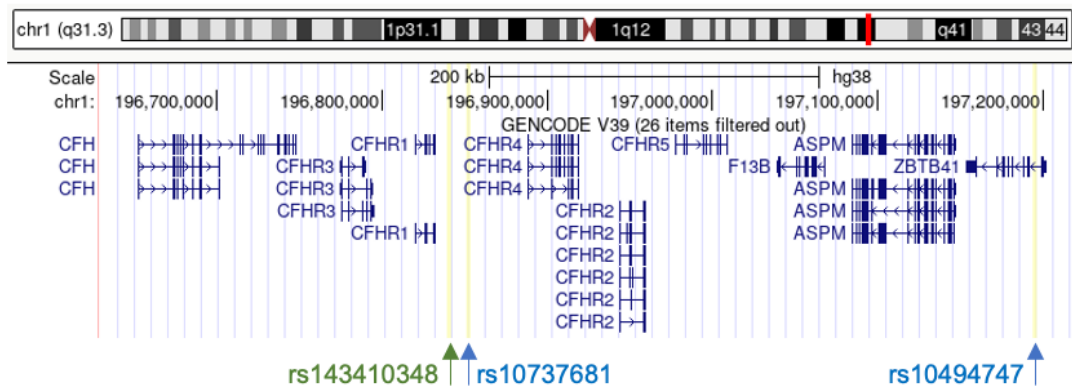
## Figure 4

medRxiv preprint doi: <https://doi.org/10.1101/2022.04.20.22274046>; this version posted April 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

### A.



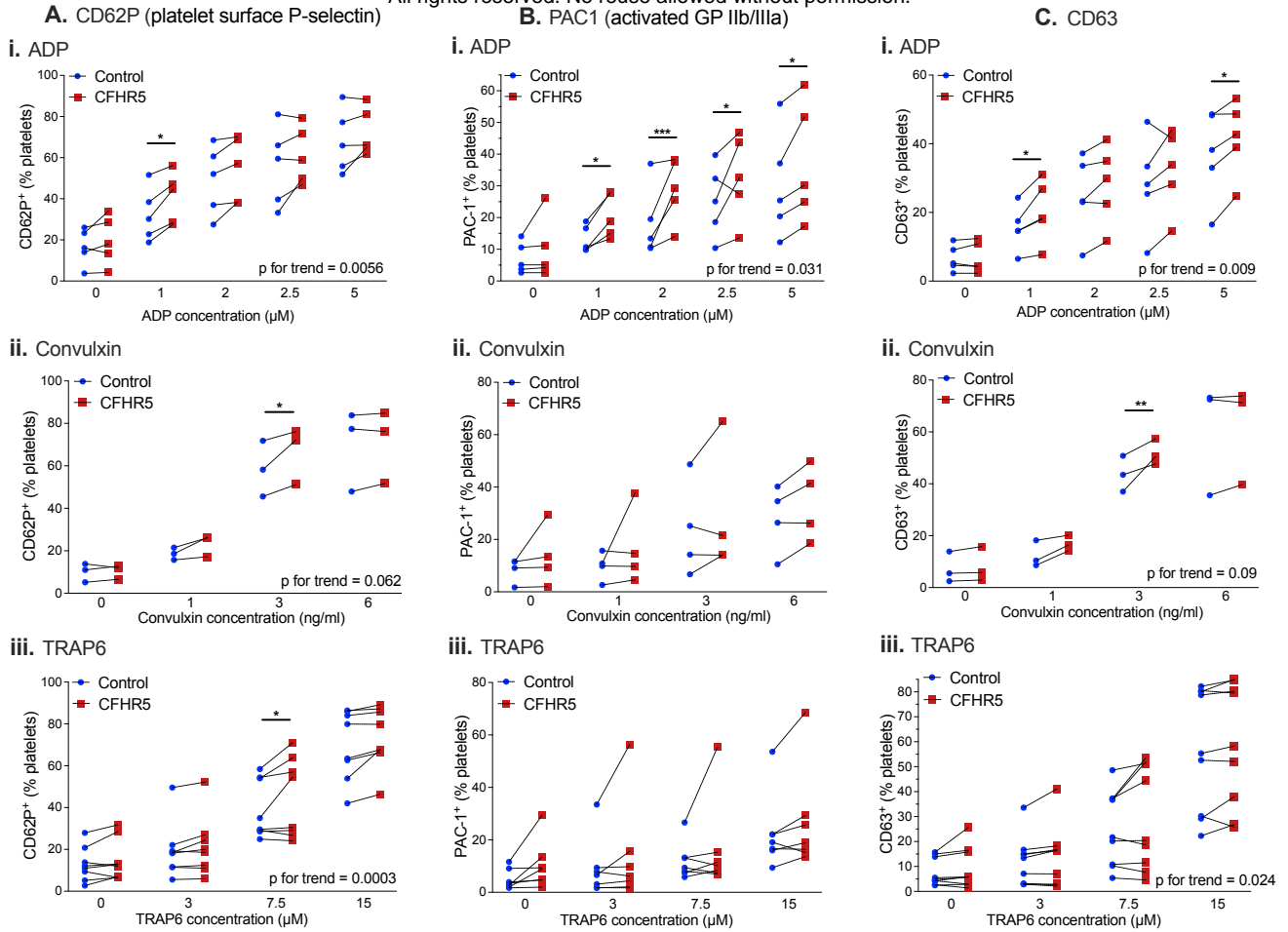
### B.



**Figure 4: GWAS analysis identifies a CFHR5 pQTL on Chromosome 1 q31.3** A meta-analysis of GWAS data for 7,135,343 SNPs tested for association with CFHR5 concentrations in a total sample of 2967 individuals from the *FARIVE*, *MARTHA* and *RETROVE* studies: **(A)** identified one genome-wide significant ( $p < 5E-08$ ) signal on chr1q31.3. **(B)** The lead SNP at this locus, rs10737681, maps between the *CFHR1* and *CFHR4* gene loci in the gene cluster of *CFHR1-5*. A borderline significant association ( $p = 9.83E-08$ ) with CFHR5 levels was observed at the rs10494747, mapping to the *ZBTB41* gene. Indicated in green is the rs143410348 recently identified with genome wide significance as associated with VTE risk [68].

## Figure 5

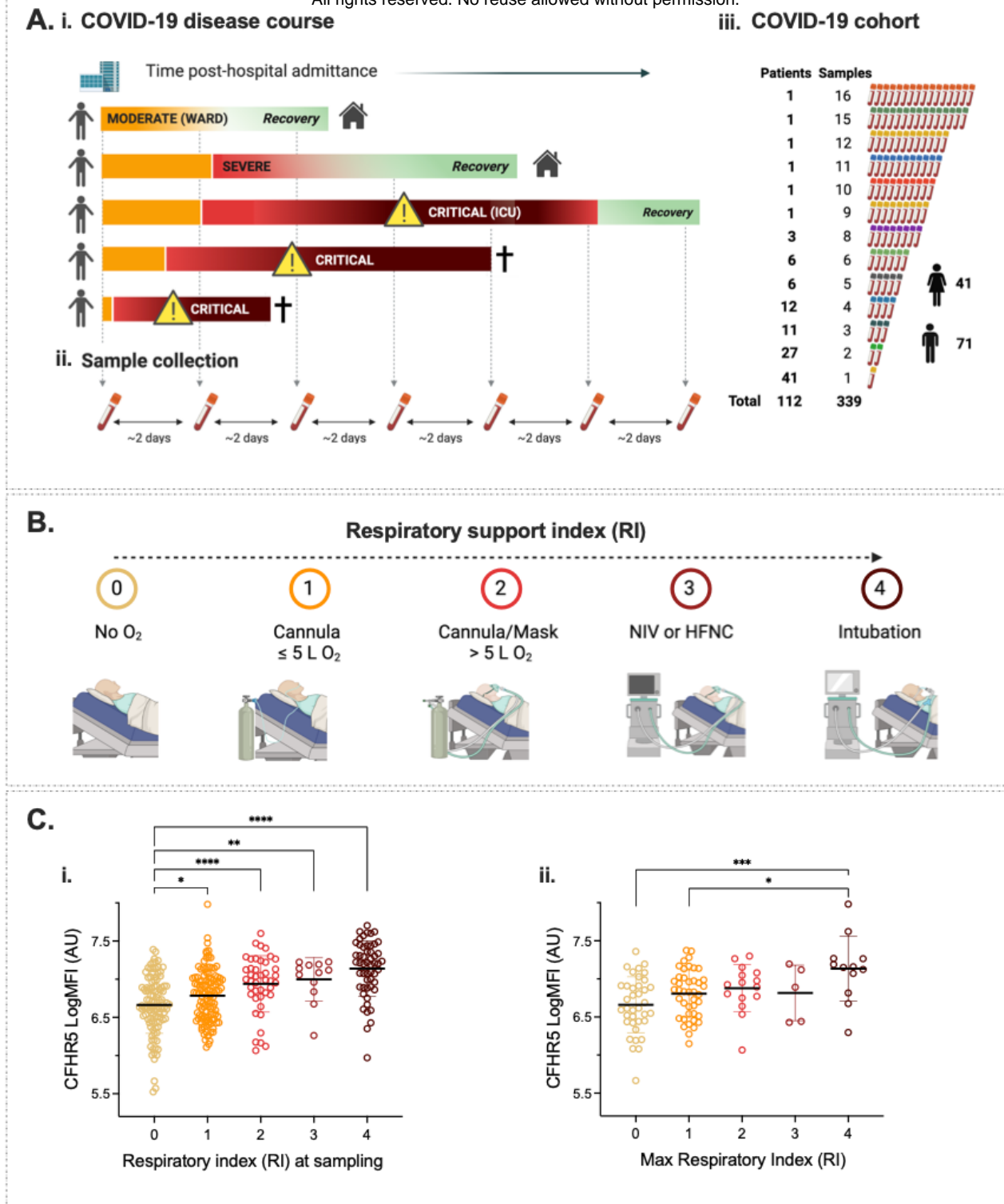
medRxiv preprint doi: <https://doi.org/10.1101/2022.04.20.22274046>; this version posted April 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



**Figure 5: Recombinant CFHR5 enhances platelet activation in platelet rich plasma.** Platelet activation was measured by surface expression of (A) P-selectin, (B) activated GP IIb/IIIa or (C) CD63, following treatment of platelet rich plasma with different concentrations of (i) adenosine diphosphate (ADP) (ii) convulxin or (iii) TRAP6, following pre-incubation (10 minutes) with recombinant CFHR5, or PBS control. Each experiment is represented by an individual point and paired experiments connected by a dotted line. \* $p < 0.05$  \*\* $p < 0.01$  \*\*\* $p < 0.001$  (p for trend bottom right).

Figure 6

medRxiv preprint doi: <https://doi.org/10.1101/2022.04.20.22274046>; this version posted April 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



**Figure 6. Plasma CFHR5 concentration is associated with COVID-19 severity**

(A) Schematic representation of the *COMMUNITY* cohort; (i) patients hospitalised with acute COVID-19 were included prospectively and followed longitudinally for disease severity (ii). Plasma was sampled at inclusion and, on average, every second day during hospital stay. (iii) In total, 359

severity registered at time of sampling was classified into a Respiratory support Index (RI): RI=0: no Oxygen (O<sub>2</sub>), RI=1: ≤ 5 litres (L) of O<sub>2</sub> administered via cannula, RI=2: ≥ 5 L O<sub>2</sub> via cannula or face mask, RI=3: Oxygenated air via Non-Invasive Ventilation (NIV) or High Flow Nasal Cannula (HFNC), RI=4: Intubated on ventilator. (C). CFHR5 plasma concentrations were measured in (i) all 339 samples to analyse the relationship with respiratory support index (RI) recorded at the time of sampling or (ii) in the 112 baseline samples to analyse the relationship with the maximum RI for that patient recorded at any subsequent time point (ii). \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

**Table 1.** Clinical characteristics of the VEBIOS ER sample set.

Variables	Cases (n=48)		Controls (n=48)	
	n	Frequency	n	Frequency
<i>Thrombosis localisation</i>				
DVT, lower limbs	20	0.42	-	-
- Proximal	17	0.85	-	-
PE	21	0.44	-	-
DVT and PE	7	0.15	-	-
<i>Sex and biometry</i>				
Sex; women	22	0.45	22	0.45
Age (years) (mean ± SD and range)	56.6 ± 18.8	(19-89)	56.6 ± 16.8	(23-88)
Current smoking ‡	4	0.08	9	0.19
<i>Family history</i>				
VTE, First degree relative < 60 years old	14	0.29	6	0.12
<i>Provoked risk factors (all)†</i>				
Estrogen containing contraceptives and hormone replacement therapy*	5	0.10	5	0.10
<i>Concentration of markers measured</i>				
CFHR5 (ng/mL)	Median	IQR	Median	IQR
	3361	(2997-3761)	2680	(2482-3075)
C3 (µg/mL)	653	(524-805)	656	(516-795)
D-dimer (ng/mL)	3604	(1335-7068)	521	(305-1128)
CRP (mg/L)	18	(8.5-42)	4.5	(2-21)
LPK (x10 <sup>9</sup> /L)	9.1	(7.2-10.2)	8	(7-10)
Hb (g/L)	135	(128-144)	138	(131-145)
TPK (x10 <sup>9</sup> /L)	206	(171-265)	240	(207-298)

**DVT**, deep vein thrombosis; **PE**, pulmonary embolism; **n**, numbers; **SD**, standard deviation; †, within one month from diagnosis or index date (immobilization with trauma, surgery, cast and/or orthosis and bedrest more than 3 days of sickness); ‡, within the last year; \*, on-going treatment; **IQR**, interquartile range; **CFHR5**, Complement Factor H-related protein 5; **C3**, Complement 3; **CRP**, C-reactive protein; **LPK**, leucocytes; **Hb**, haemoglobin; **TPK**, thrombocytes. Missing values across all variables are <10%.



**Table 2.** Risk of VTE associated with CFHR5 concentration in five independent studies.

		VEBIOS ER	DFW-VTE	FARIVE	VEBIOS Coagulation	RETROVE
<b>VTE ALL</b>	OR (1SD)	2.54	1.80	1.24	1.55	1.29
	(95%CI)	(1.52-4.66)	(1.29-2.58)	(1.10-1.40)	(1.2-2.01)	(1.09-1.53)
	P-value	1.05E-03	7.65E-04	3.98E-04	8.85E-04	2.4E-03
	Cases	48	54	582	142	308
	Controls	48	146	576	135	360
CFHR5 range (ng/mL)		1216-6073	232-1170	450-4904	1378-4570	364-2341
<b>Tertile 1</b>	OR (1SD)	Ref.	Ref.	Ref.	Ref.	Ref.
	(95%CI)					
	P-value	NA	NA	NA	NA	NA
	Cases	10	9	165	39	91
	Controls	22	57	221	54	132
CFHR5 range (ng/mL)		1216-2653	232-476	450-1315	1378-2333	364-858
<b>Tertile 2</b>	OR (1SD)	1.64	2.72	1.40	1.41	1.03
	(95%CI)	(0.58-4.80)	(1.15-6.86)	(1.05-1.86)	(0.78-2.55)	(0.69-1.52)
	P-value	0.36	0.03	0.02	0.25	0.88
	Cases	13	21	198	45	94
	Controls	19	46	188	47	128
CFHR5 range (ng/mL)		2675-3338	478-601	1317-1716	2335-2779	859-1063
<b>Tertile 3</b>	OR (1SD)	9.05	2.93	1.75	2.51	1.67
	(95%CI)	(2.93-31.61)	(1.24-7.37)	(1.31-2.33)	(1.39-4.63)	(1.13-2.47)
	P-value	2.51E-04	0.02	1.28E-04	2.68E-03	9.90E-03
	Cases	25	24	219	58	123
	Controls	7	43	167	34	100
CFHR5 range (ng/mL)		3345-6073	603-1170	1717-4904	2782-4569	1064-2341