THE LANCET
Digital Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Kotanidis CP, Xie C, Alexander D, et al. Constructing custom-made radiotranscriptomic signatures of vascular inflammation from routine CT angiograms: a prospective outcomes validation study in COVID-19. *Lancet Digit Health* 2022; published online Aug 26. https://doi.org/10.1016/S2589-7500(22)00132-7.

- Online Methods -

Inclusion & Exclusion criteria

Patients in Study Arm 1 originated from the AdipoRedOx study (Oxford REC C: 11/SC/0140), which enrols patients undergoing cardiac surgery [including coronary artery bypass grafting (CABG) and valve replacement/repair] at the John Radcliffe hospital, Oxford University NHS Foundation Trust, UK. Exclusion criteria include inflammatory, neoplastic, renal, or hepatic diseases. All subjects in Study Arm 1 have given written informed consent for sample and data collection, and follow-up imaging. Subjects in Study Arm 2 are part of the prospective arm of the Oxford Risk Factors and cArdiovascular imagiNg (ORFAN) study (Oxford REC C: 15/SC/0545, NCT05169333). These patients gave written informed consent to participate in the study and came back for a follow up CCTA scan. The patients in Study Arms 3 and 4 are part of the registry arm of the Oxford Risk Factors and cArdiovascular imagiNg (ORFAN) study (Oxford REC C: 15/SC/0545, NCT05169333), and the collection of pseudoanonymised data was performed under Section 251 (NHS Act 2006), with specific approval from the Confidentiality Advisory Group (CAG, reference 20/CAG/0157), as defined in the ORFAN Study protocol.

Severity of COVID-19

The severity of COVID-19 infection in our population was defined according to the WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection scoring system for hospitalised patients, as follows: mild disease: hospitalised patients not requiring oxygen therapy (score 4); moderate disease: patients requiring oxygen by mask or nasal prongs

(score 5); severe disease: patients supported with non-invasive ventilation (score 6); critical: patients supported with intubation and mechanical ventilation (score 7-9)[1].

Tissue Collection, RNA Isolation and Sequencing Library Preparation

In Study Arm 1, IMA specimens were collected during surgery and stored in TRI reagent (Sigma, catalogue number T9424) at -80^{o}C until thawed for RNA isolation. Total RNA was isolated by a phenol to chloroform (1:5 ratio) separation protocol followed by a magnetic beads-based RNA purification method on a KingFisher magnetic particle processor (Thermo Fisher Scientific), using the MagMAX mirVana total RNA isolation kit (Thermo Fisher Scientific, Catalogue Number A27828). RNA concentration was assessed spectrophotometrically on NanoDrop ND-1000. For RNA sequencing, the QuantSeq 3' mRNA (Lexogen) library preparation kit was used. All samples were sequenced as part of a large multiplex pool on an Illumina NovaSeq 6000 system producing 150bp paired-end reads. The COMBAT whole blood RNAseq dataset was generated as described previously[2]. Briefly, whole blood was collected into Tempus tubes (Life Technologies) and frozen at -80^{o}C until extraction in batches. Total RNA-seq was performed with libraries prepared by Oxford Genomics Centre with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina after rRNA and globin depletion. Libraries were sequenced as a single pool of 144 samples (124 patients) on one NovaSeq S4 flow cell (4 lanes) with a target of 50M 100bp read pairs per sample.

Sequencing Data Processing and Analysis

RNAseq read pairs were split and only read 1 was used for alignment and quantification, appropriate for the library preparation protocol used. After initial poor mapping results with full-

length 150bp reads, reads reduced to 75bp and trimmed for adapter and low-complexity sequence using fqtrim v0.9.5[3] were aligned to Homo sapiens reference genome, GRCh37 using Hisat2 version-2.0.4[4]. Gene annotation files were downloaded in GTF format from Ensembl, release 76[5]. Reads mapping to annotated exon features were quantified with featureCounts[6], part of subread-v1.5.0[7], using default parameters and keeping all aligned reads including potential duplicates (due to the 3' protocol). Quality metrics for aligned reads were estimated using CollectRnaSeqMetrics, implemented in Picard tools v1.92[8]. The raw count tables produced by featureCounts were imported in R Statistical Software[9] for further processing and analysis using in-built functionality and relevant packages as detailed below. In COMBAT, adaptor sequences were removed with TrimGalore, reads aligned to the reference genome (GRCh18) using STAR and read counts generated with featureCounts and annotations from Ensembl (v100). One poor quality sample was removed, and features filtered based on a threshold of >10 reads in >10 samples. The data were normalised using the trimmed mean of M-values method from edgeR and log2-transformation. Samples (n=30) from the 23 patients with a CT scan and therefore C19-RS calculated were extracted from this complete dataset.

Clustering and Differential Expression Analysis

The set of inflammatory genes was extracted from the count table and filtered to exclude those with low expression levels: counts per million (CPM) values were generated using the edgeR package[10] and 51 genes with CPM > 1 in at least 10 out of 55 patients were retained. Unsupervised hierarchical clustering was performed on this set of genes with the Minkowksi distance metric (p=10) and ward.D clustering method (heatmap.2 function in the gplots package[11]) and visualised as a heatmap with dendrogram. The 2 clusters of samples identified based on the inflammatory

expression profile were then assessed for differential expression (edgeR package[10]) on the full set of expressed genes in the RNA-Seq dataset. This generated a list of 132 genes identified as differentially upregulated (logFC > 1 and $p < 0.05$) between the 2 clusters of patients. In order to validate our clusters for activation of inflammation within the vasculature we conducted pathway enrichment analysis in ConsensusPathDB-human with the up-regulated differentially expressed genes (DEGs); the input for pathway analysis was the set of 132 genes with logFC > 1 and $p < 0.05$.

In the COMBAT dataset, 77 of the inflammatory genes had been retained post-filtering on the full dataset. The scaled expression of these genes across the range of C19-RS were visualised as a heatmap using the pheatmap R package, and unsupervised clustering was performed as above. Differential expression analysis was performed on protein-coding genes using one sample per patient (closest timepoint to CT scan) and the limma package, comparing high (above median) and low (below median) C19-RS scores and adjusting for age and sex. Multiple testing correction was performed using the Benjamini-Hochberg method.

Weighted Gene Co-expression Network Analysis (WGCNA) was applied through the "cornet" pipeline to identify modules of correlated genes in the complete COMBAT dataset (143 samples from 123 patients) (Langfelder and Horvath, 2008, https://github.com/sansomlab/cornet.git). In brief, this uses a stepwise approach of correlation network construction and module detection. Using the soft thresholding power of 4, a signed-hybrid network was built, with the biweight midcorrelation as the adjacency function. The adjacency matrix was transformed into a topological overlap matrix to calculate the dissimilarity, and a dissimilarity threshold of 0.3 was used to merge

modules with very similar expression profiles. Module eigengene values (module first principal component) were used to summarise modules and perform module-C19-RS correlation analysis (Pearson correlation). Pathway enrichment analysis was performed using the hypergeometric test with the default settings of the "cornet" pipeline (https://github.com/sansomlab/cornet.git).

Viral Genome Sequencing

Samples were sequenced using a multiplex PCR based approach with the ARTIC LoCost protocol[12] and v3 primers1using R9.4.1 flow cells (Oxford Nanopore Technologies, Oxford, UK). Consensus sequences were generated using ARTIC field bioinformatics v1.2.1.[13] All sequences underwent quality control, requiring >90% consensus genome coverage at ≥20 depth. Lineages were assigned with Pangolin[14].

Coronary and Pulmonary Computed Tomography Angiography Imaging and Acquisition Protocols

In Study Arm 1 participants underwent coronary CTA (CCTA) prospectively after patient consent using a 64-slice scanner (LightSpeed Ultra or Revolution GSI, General Electric) as previously described[8]. Heart rate was optimised using intravenous injection of beta-blockers and sublingual glyceryl-trinitrate (800ug) was also administered to achieve maximum coronary vasodilatation. CCTA was performed following intravenous injection of 95ml of iodine-based contrast medium (Niopam 370, BRACCO) at a flow rate of 6mL/sec (tube energy of 120 or 100 kVp, axial slice thickness of 0.625 mm, rotation time of 0.35 sec, detector coverage of 40 mm). Prospective image acquisition was used by ECG-gating at 75% of cardiac cycle (with 100 msec padding for optimal imaging of the right coronary artery if required). In Study Arm 2, we identified patients with

existing CCTA imaging, who were recruited prospectively into the ORFAN study before the pandemic and were participating into the prospective follow up part of the study (NCT05169333). Within that cohort, we then identified patients who had confirmed COVID-19 infection in the past 6 months, and we matched them 1:1 for age, sex, and BMI to control cases from the same cohort who had never been diagnosed with COVID-19 until the time of the screening. These patients were then invited to return for their follow up research CCTA scan, as part of the ORFAN study protocol, which was done prospectively, maintaining identical scanning settings as for their baseline scan. CCTA scans were performed on a 320-slice scanner (Aquilion One, Canon Medical Systems, Tochigi, Japan) In patients with heart rate > 65 beats/minute, 5 mg of intravenous metoprolol (with incremental 5 mg doses up to a maximum dose of 40 mg) Patients also received 0.8 mg of nitroglycerin sublingually immediately before CCTA and iodinated contrast (Iomeron 350, Bracco UK Ltd) was administered at flow rate of 5-6 ml/s. In Study Arms 3 and 4, we identified the consecutive patients who were hospitalised with COVID-19 and had a clinically-indicated CTA of their pulmonary arteries at the Oxford University Hospitals NHS Trust, University Hospitals of Leicester NHS Trust and Royal United Hospitals Bath NHS Trust between March 2020 - January 2021. These were existing scans that were analysed blindly by two operators within the quality management System of the Oxford Cardiovascular CT (OXACCT) core lab. In Study Arm 3, participants underwent pulmonary CTA (CTPA) on the GE Revolution HD CT scanner. During the initial phase of the pandemic, suspected cases of COVID-19 pneumonia had standard CTPA or dual energy CTPA (DECTPA)[15]. The single energy CTPA were perform using 80-120kV and contrast 70-100mL at 4 mL/s dependent on body size. DECTPA were performed using rapid kV switching to optimise contrast and thrombus visualisation. These scans were all non-ECG gated. In Study Arm 4, scans from University Hospitals of Leicester were performed on

a Siemens Somatom Definition Flash CT scanner, using a FLASH protocol with the following scan parameters: dose modulation on, quality reference kVp 100, ref mAs 100, pitch: 2.1, rotation time 0.28s, detector configuration 128x0.6mm, suspended respiration scan from lung apices to lung bases, using a pre monitoring slice at the level of the pulmonary artery for optimal contrast enhancement of the pulmonary arteries, contrast agent: omnipaque 350, 50mls at 4ml/sec plus 50mls saline flush. Scans from the Royal United Hospitals Bath were performed on either Siemens Somatom Definition Edge scanner with the following scan parameters: dose modulation on, quality reference kVp 120, ref mAs 145, pitch 1.2, rotation time 0.5s, detector configuration 128x0.6mm, suspended respiration scan from lung apices to lung bases, using a bolus-tracking method with threshold of 100 HU in an ROI in a slice at the level of the main pulmonary artery and a 4 second delay after triggering, contrast agent: omnipaque 350, 60mls at 5ml/s plus 50mls 0.9% saline flush at 5ml/sec or Siemens Somatom Drive scanner with the following scan parameters: dose modulation on, dual source dual energy 80 kVp / ref mAs 141 and Sn-filter 140 kVp / ref mAs 60, pitch 1.2, rotation time 0.5s, detector configuration 128x0.6mm, suspended respiration scan from lung apices to lung bases, using a bolus-tracking method with threshold of 100 HU in an ROI in a slice at the level of the main pulmonary artery and a 9 second delay after triggering, contrast agent: omnipaque 350, 75mls at 5ml/s plus 25mls 50:50 mix omnipaque 350 / 0.9% saline flush at 5ml/sec.

Adipose Tissue Segmentation and Radiomic Characterization

Prior to segmentation, all scans were screened for image quality, and ones deemed as poor quality were excluded from further analysis (**appendix p 18**). Image processing and extraction of radiomic features was performed using the CaRi-Research™ toolbox 2.1.1. (Caristo Diagnostics, Oxford

UK).[16] Perivascular adipose tissue segmentation was performed manually around the right IMA from the level of the aortic arch to 120mm caudally. Perivascular space was defined as the space within a radial distance from the outer vessel wall equal to the diameter of the respective vessel, as previously described[17]. A segmentation tool was used to track a cylindrical segmentation area around the internal mammary artery with a diameter as described above. Following this, manual corrections were made, so as to exclude any lung tissue or any other type of tissue posterior to the intrathoracic fascia, which would not be in direct contact with the internal mammary artery. Following this step, the segmentation was computationally thresholded to an attenuation window of -190 to -30 HU in order to isolate perivascular adipose tissue only. Similarly, perivascular adipose tissue segmentation around the descending thoracic aorta was performed manually from the level of the pulmonary artery bifurcation to 67.5mm caudally, as previously described[18]. Perivascular space was defined as the space within a cylindrical layer that is expanded beyond the vessel borders by a distance equal to 10mm. In order to avoid lung tissue and any COVID-19 related lesions, peri-aortic adipose tissue directly adjacent to the left lateral side of the descending thoracic aorta was removed. Again, the segmentation was computationally thresholded to an attenuation window of -190 to -30 HU in order to isolate perivascular adipose tissue only. The segmentations were performed by two experienced researchers, according to a Standard Operating Procedure (SOP), developed within the Oxford Academic Cardiovascular CT Core lab. This SOP provides clear instructions on how the segmentations are performed, and it includes specific process for training of the operators, and criteria to sign them off as competent to perform this analysis. This SOP is part of the Quality Management System of the OXACCT core lab (version 0.1, May 2020), which is a qualified imaging core-lab supporting academic as well as industrial clinical trials. For this study, CPK and CX were the two clinically qualified operators who

performed the analyses of the scans, blinded to group allocation. For C19-RS measurements inter-observer agreement amongst two independent operators was excellent ICC: 0.914. To address this further, in post-hoc analyses we added the rater as a covariate in our multivariable models, and that had no impact on the significance or effect size of C19-RS on the prediction of in-hospital outcomes. Specifically, the HR for C19-RS in predicting in-hospital mortality was 3.20 [95%CI: 1.42-7.19], p=0.005 in Arm 3 and 2.85 [95%CI 1.20-6.75], p=0.01 in Arm 4, when the rater was included into the model together with age, sex, cardiovascular risk factors (hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease), C-reactive protein plasma levels, white blood cell count, plasma troponin, history of chronic obstructive pulmonary disease and CT tube voltage. Given the heterogeneity of tube voltage and effective energies used in the various imaging protocols, we rescaled all images using 100kVp as reference. Conversion factors (**appendix p 39**) for tube voltage 120kVp and effective energies 55, 58, and 70 keVs were calculated from data previously validated[19]. Radiomic features were extracted using CaRi-Research^{TM} 2.1.1. toolbox (Caristo Diagnostics, Oxford UK) and pyradiomics[16].

A total of 1,655 radiomic features were extracted from each segmented PVAT volume. Briefly, the radiomic features obtained were based on shape-based analysis, first order statistics, grey level co-occurrence matrix (glcm), grey level run length matrix (glrlm), grey level size zone matrix (glszm), grey level dependence matrix (gldm), neighbouring grey tone difference matrix (ngtdm). All these features were calculated on the original images and after applying transformations on the images. The transformations included Laplacian of Gaussian (log) with various sigma (1, 2, 3, 4, 5 mm), wavelet-LLH, wavelet-LHL, wavelet-LHH, wavelet-HLL, wavelet-HLH, wavelet-HHL, square, square root, logarithm and exponential. The bin width was kept at 25.

Radiomic feature filtering and XGBoost modelling

Firstly, we performed a stability assessment of all 3,310 different radiomic features. For this purpose we used 24 Lung CT scans (12 paired scans, performed 15 minutes apart) from the RIDER dataset[20] to assess the scan-rescan ICC of each radiomic feature. Only radiomic features with ICC≥0.90 were included in further analyses (n=2,177, **appendix p 27**). Although the RIDER dataset was not designed with inflammation in mind, it still provides a valid technical dataset that allows reliable testing of the reproducibility of radiomic features in the PVAT around human IMAs. We next filtered out redundant radiomic features with Spearman's rho coefficient lower than an absolute value of 0.9, using the "findCorrelation" function of the "caret" package in R (n=497). For further filtering of radiomic features we removed features significantly correlated with BMI and total intrathoracic adipose tissue, using a threshold of 0.05 in the Spearman's rho correlation p value (n=333). Next, we isolated a randomly split 20% exploratory subset from the Study Arm 3 population, applied a univariate ROC analysis for outcome prediction in that subset (outcome: COVID-19 positive status) as well as in Study Arm 1 (outcome: high vascular inflammation), and filtered in only those radiomic features that predicted the outcome in both datasets in the same direction (n=144). Finally, recursive feature elimination with a random forest algorithm and repeated five-fold cross-validation showed a plateau in the accuracy of the trained model with a selection of 33 final features (**appendix p 28**).

Radiomic features that were retained after filtering, were scaled and fit in an extreme gradient boosting algorithm. We considered a series of machine learning methodologies to use in our datasets, including other decision trees, and random forest algorithms (**appendix p 40**). Extreme

gradient boosting with method "gbtree" was the method with the best performance. We therefore decided to choose extreme gradient boosting, also because this is the de facto standard algorithm for getting accurate results from predictive modelling with machine learning. It's the fastest gradient-boosting library with very high accuracy. It is understood to perform well in many applications with a fair amount of data and can detect and learn from non-linear data patterns[21]. The final product of the XGBoost algorithm (namely the raw logit values) was defined as C19-RS. To avoid overfitting issue, we used 5-fold cross validation and tuned the hyperparameters by optimising step size shrinkage, L2 regularization parameter and learning rate. We used early stopping technique that works by monitoring the performance of the model that is being trained on a separate test dataset and stopping the training procedure once the performance on the test dataset has not improved after a fixed number of training iterations. It avoids overfitting by attempting to automatically select the inflection point where performance on the test dataset starts to decrease while performance on the training dataset continues to improve as the model starts to overfit. Although modest number of samples were used during training step of model development, the accuracy of the model was not degraded on external validation data suggesting that model was not overfitted.

Statistical Analysis

Participant demographics are summarized as numbers (percentages) or median (25th to 75th percentile) for categorical and continuous variables, respectively. Between-group comparisons were performed using Pearson's chi-squared test for categorical variables, and the Mann-Whitney independent samples test for numeric variables. As for power calculations[22], Study Arm 1 was exploratory and served as the development set for C19-RS. In Study Arm 2, power calculations

were performed in advance, in order to define the sample size needed to recall for a follow up scan. We calculated that for COVID-19 positive patients in order to explore a delta in C19-RS values between baseline and follow-up scans of 1.7 with standard deviation of 2.3, we would require a sample size of 21. This sample size would offer us statistical power 0.9 to detect a difference of 1.7 arbitrary units (AU) in delta(C19-RS) between cases and controls, for SD 2.3 and α=0.05. In Study Arm 3, we calculated that for a population of 250, of which a third would have high C19-RS values, we would be able to detect a minimum hazard ratio of 1.64 with power 0.9 and α=0.05. For Study Arm 4, we calculated that for a hazard ratio of 3.31 taken from the internal test cohort, power 0.8 and alpha 0.05, we would require 22 events. Missing values within the datasets were imputed with predictive mean matching using the "mice" package. We further performed post-hoc exploratory subgroup analyses excluding variables with missingness greater than 10%. Statistical analyses were performed in the R environment (R version 3.6.0 and R Studio version 1.2.1335; COMBAT whole blood: R version 3.6.2, RStudio 1.2.5042 running on the BMRC compute cluster). All tests were two-sided and α was set at 0.05. When C19-RS was measured, the outcomes data were collected and the statistical analysis took place as post-hoc investigation of prospectively collected data. Model development and reporting followed TRIPOD (transparent reporting of a multivariable prediction model for individual prediction or diagnosis) guidelines (**appendix p76**)[23].

Study Arm 1

Hierarchical clustering was performed using Ward's method and Minkowski distance with p set to 10. Radiomic features that were retained after filtering (described above), were scaled and fit in an extreme gradient boosting algorithm. The method chosen was decision trees, with eta set to 0.5,

number of rounds 100, and maximum tree depth 50. The final product of the XGBoost algorithm was defined as C19-RS.

Study Arm 2

Propensity score matching for age, sex, and BMI was performed to compare 22 COVID-19 patients with serial CCTA images and non-COVID-19 controls. Wilcoxon signed-rank test was used for paired comparisons, and unpaired Mann-Whitney U test was used for between groups comparisons.

Study Arm 3

In the overall Study Arm 3 population we first assessed C19-RS' ability to detect COVID-19 in multivariable logistic regression. An optimal cut-off point for C19-RS was determined by the Youden's statistic method in receiver operating characteristic curve analysis. In the COVID positive population only (n=254) C19-RS' prognostic value for in-hospital death "due to" or "involving" COVID-19 as defined by the Office of National Statistics (ONS), and a composite endpoint of in-hospital death and intensive care unit (ICU) admission was assessed in receiver operating characteristic curves, univariate Kaplan–Meier curves, and logistic and Cox regression models adjusted for age above 65, sex, cardiovascular risk factors (hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease), C-reactive protein plasma levels, white blood cell count, plasma troponin, history of chronic obstructive pulmonary disease and CT tube voltage. The optimal C19-RS cut-point for survival analyses was identified by the value that maximized the log-rank statistic for death in hospital. Correlation of C19-RS with C-reactive protein and length of hospital stay was assessed using Spearman's rho. Missingness was 29.7% for BMI,

10.4% for CRP, 54.4% for troponin and below 10% for the rest of the variables. In the COMBAT total RNAseq data, differential gene expression analysis was performed using the limma R package, co-expressed gene modules identified using WGCNA, and pathway enrichment assessed used GOBP annotations and hypergeometric test as previously described[2]. Correlation of gene modules with the C19-RS signature was quantified by Pearson's r.

Study Arm 4

C19-RS' prognostic value for in-hospital death and a composite endpoint of in-hospital death and intensive care unit (ICU) admission was assessed in Cox regression models adjusted as above. Missingness was 45.2% for BMI, 75% for Troponin, and below 10% for the rest of the variables.

Online References

1. Marshall JC, Murthy S, Diaz J, et al. A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious Diseases* 2020; **20**(8): e192-e7.

2. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* 2022.

3. Pertea G. fqtrim: v0. 9.4 release. 2015.

4. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 2015; **12**(4): 357-60.

5. Cunningham F, Amode MR, Barrell D, et al. Ensembl 2015. *Nucleic Acids Res* 2015; **43**(Database issue): D662-9.

6. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**(7): 923-30.

7. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* 2013; **41**(10): e108-e.

8. Picard Tools. version at 10.01.2016. https://broadinstitute.github.io/picard/2020).

9. Team RC. R: A language and environment for statistical computing. Vienna, Austria; 2013.

10. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**(1): 139-40.

11. Warnes GR, Bolker B, Bonebakker L, et al. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.4. 2020.

12. Quick J. nCoV-2019 sequencing protocol v3 (LoCost). 2020.

13. Loman N. Artic-network field bioinformatics: 1.2.1. 2021.

14. Pangolin software package.

15. Lang M, Som A, Mendoza DP, et al. Hypoxaemia related to COVID-19: vascular and perfusion abnormalities on dual-energy CT. *The Lancet Infectious Diseases* 2020.

16. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017; **77**(21): e104-e7.

17. Oikonomou EK, Marwan M, Desai MY, et al. Non-invasive detection of coronary inflammation using computed tomography and prediction of residual cardiovascular risk (the CRISP CT study): a post-hoc analysis of prospective outcome data. *Lancet (London, England)* 2018; **392**: 929-39.

18. Schlett CL, Massaro JM, Lehman SJ, et al. Novel measurements of periaortic adipose tissue in comparison to anthropometric measures of obesity, and abdominal adipose tissue. *International journal of obesity* 2009; **33**(2): 226-32.

19. Okayama S, Soeda T, Takami Y, et al. The influence of effective energy on computed tomography number depends on tissue characteristics in monoenergetic cardiac imaging. *Radiology research and practice* 2012; **2012**.

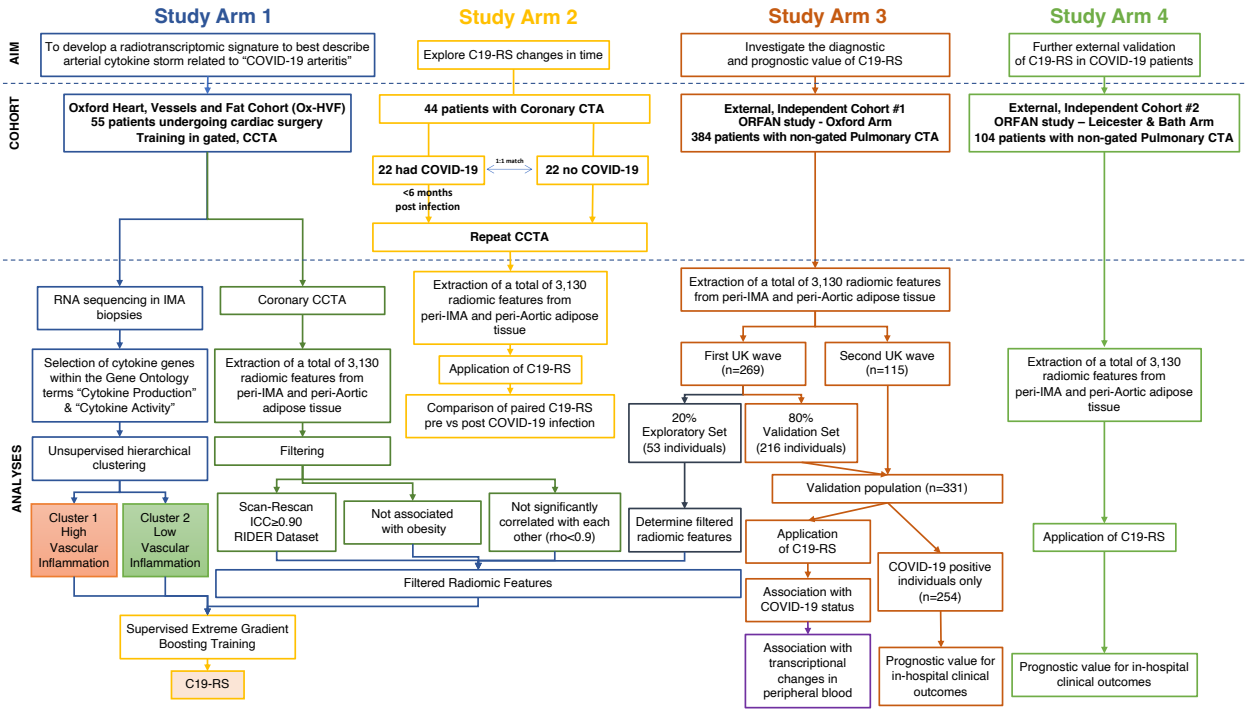
20. RIDER. The Reference Image Database to Evaluate Therapy Response. 2013.

21. Kadiyala A, Kumar A. Applications of python to evaluate the performance of decision tree-based boosting algorithms. *Environmental Progress & Sustainable Energy* 2018; **37**(2): 618-23.

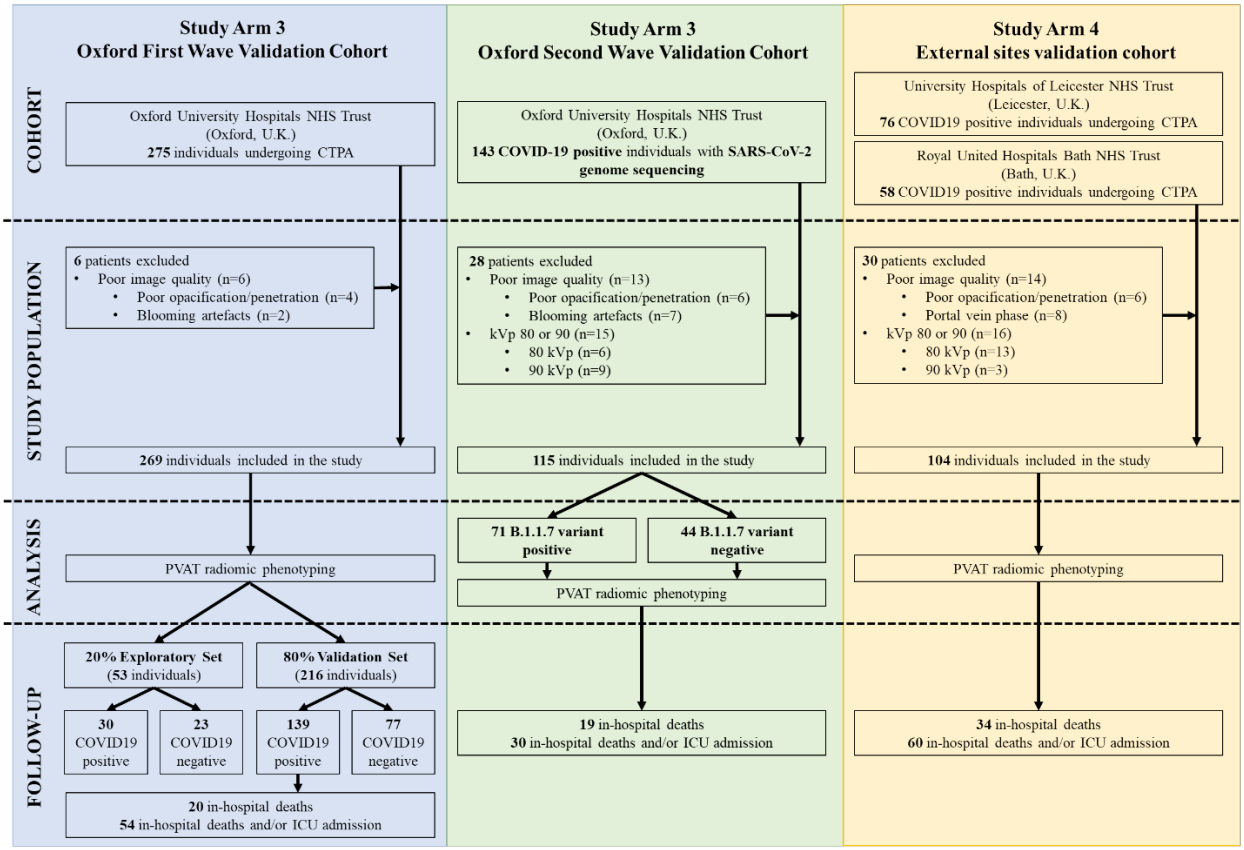
22. Dupont WD, Plummer Jr WD. Power and sample size calculations: a review and computer program. *Controlled clinical trials* 1990; **11**(2): 116-28.

23. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine* 2015; **162**(1): W1-W73.

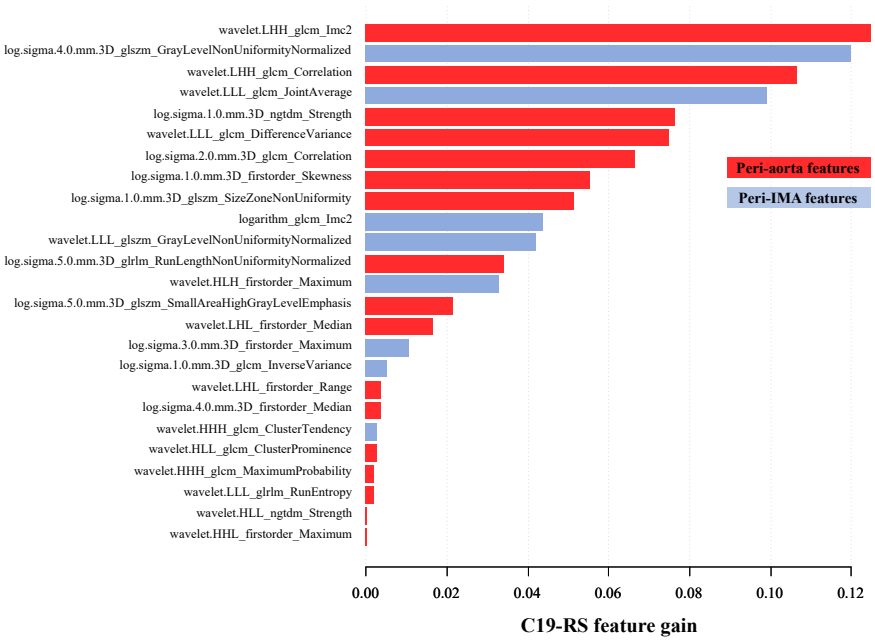
Online Figures



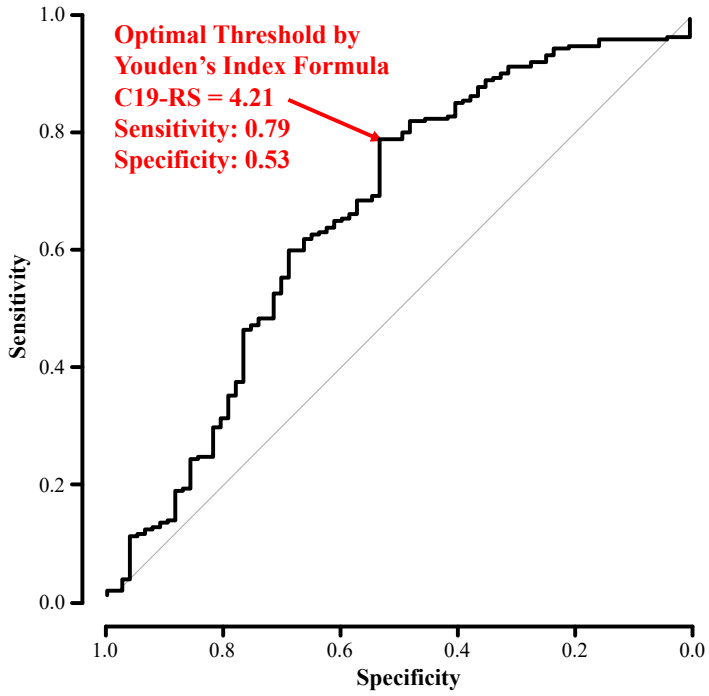
Online Figure 1 | Workflow Diagram. Study Arm 1 utilised 55 patients from the Oxford Heart Vessels and Fat (Ox-HVF) Cohort that underwent cardiac surgery and CCTA imaging in order to develop the radiomic signature C19-RS based on high vascular inflammation. Study Arm 2 included 88 paired CCTA scans. Study Arm 3 included 384 participants with CTPA imaging and was used for radiomic feature filtering and to validate C19-RS for COVID-19 discrimination (COVID-19 positive and negative individuals, n=331) and in-hospital outcome correlation (COVID-19 positive individuals only, n=254). Study Arm 4 served as an external, independent validation cohort of COVID-19 patients testing the prognostic value of C19-RS for in-hospital outcomes.



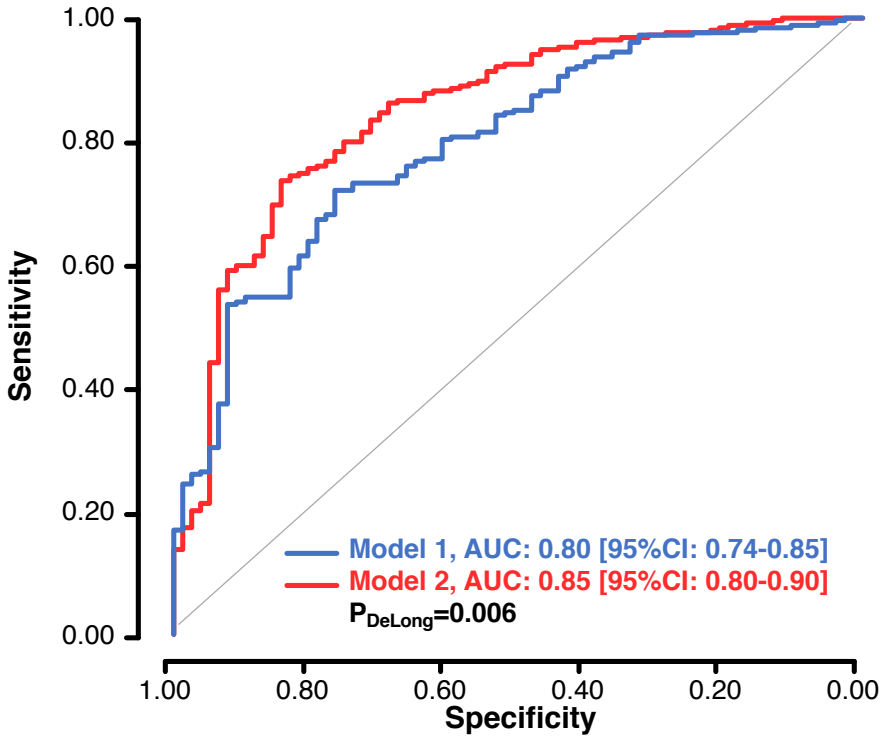
Online Figure 2 | Image analysis and in-hospital outcomes collection diagram for Study Arms 3 & 4. ICU: Intensive care unit.



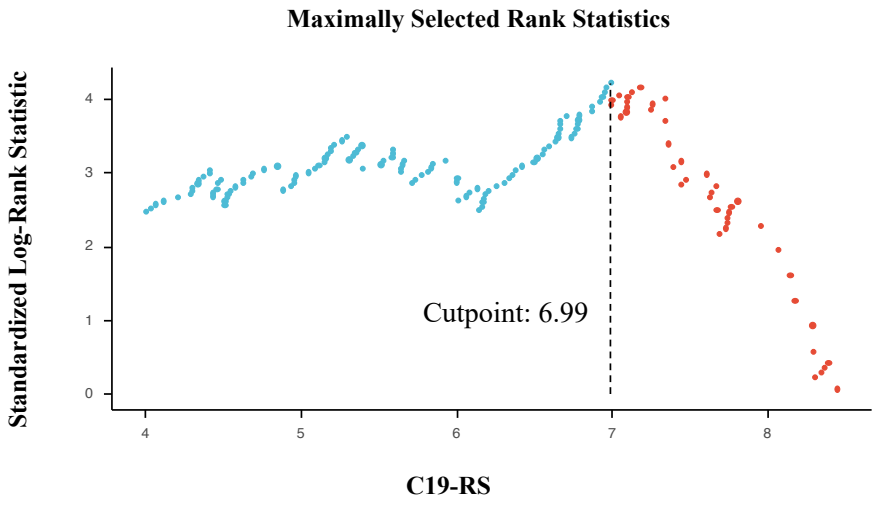
Online Figure 3 | Features gain in C19-RS. C19-RS consists of 25 features, 8 peri-IMA radiomic and 17 peri-aorta features. Gain values represent the relative contribution of each radiomic feature to C19-RS. A higher gain value when compared to another feature implies higher importance for generating a prediction value. A full list of radiomic features is presented in **appendix p 39**.



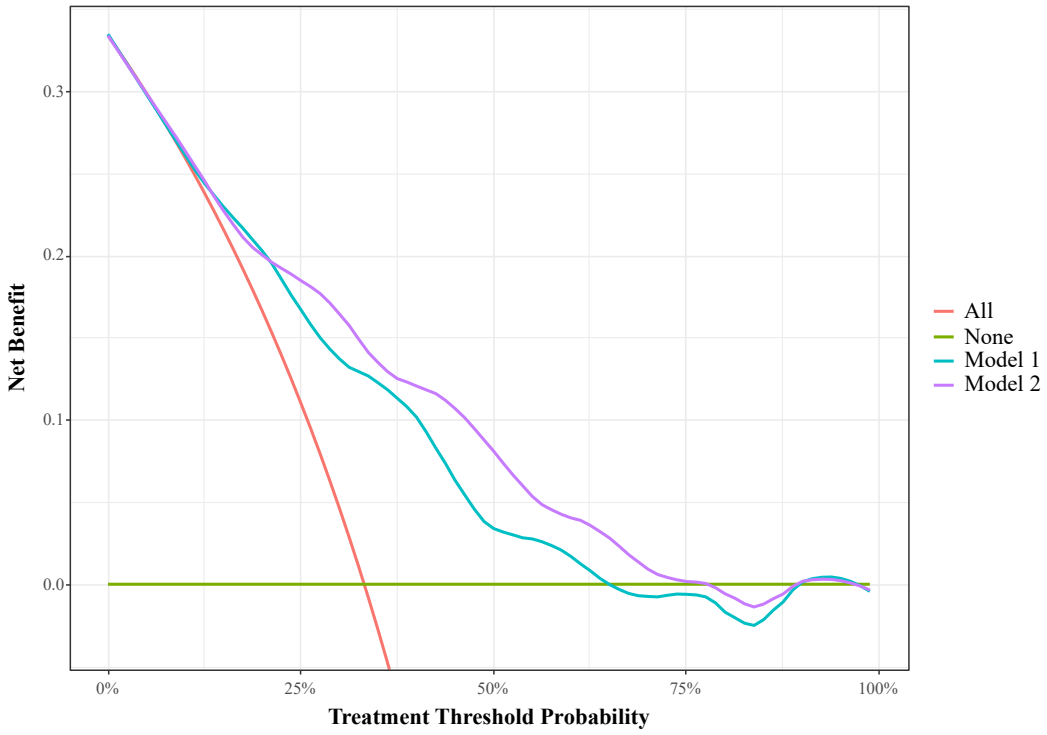
Online Figure 4 | Optimal C19-RS cut-off for COVID-19 detection. Receiver operating characteristic curve, area under the curve (0.65 [95%CI: 0.57-0.73], p<0.001) and optimal cut-off point of C19-RS for COVID-19 detection in COVID-19 negative and positive patients from the validation Study Arm 3 population (n=331).



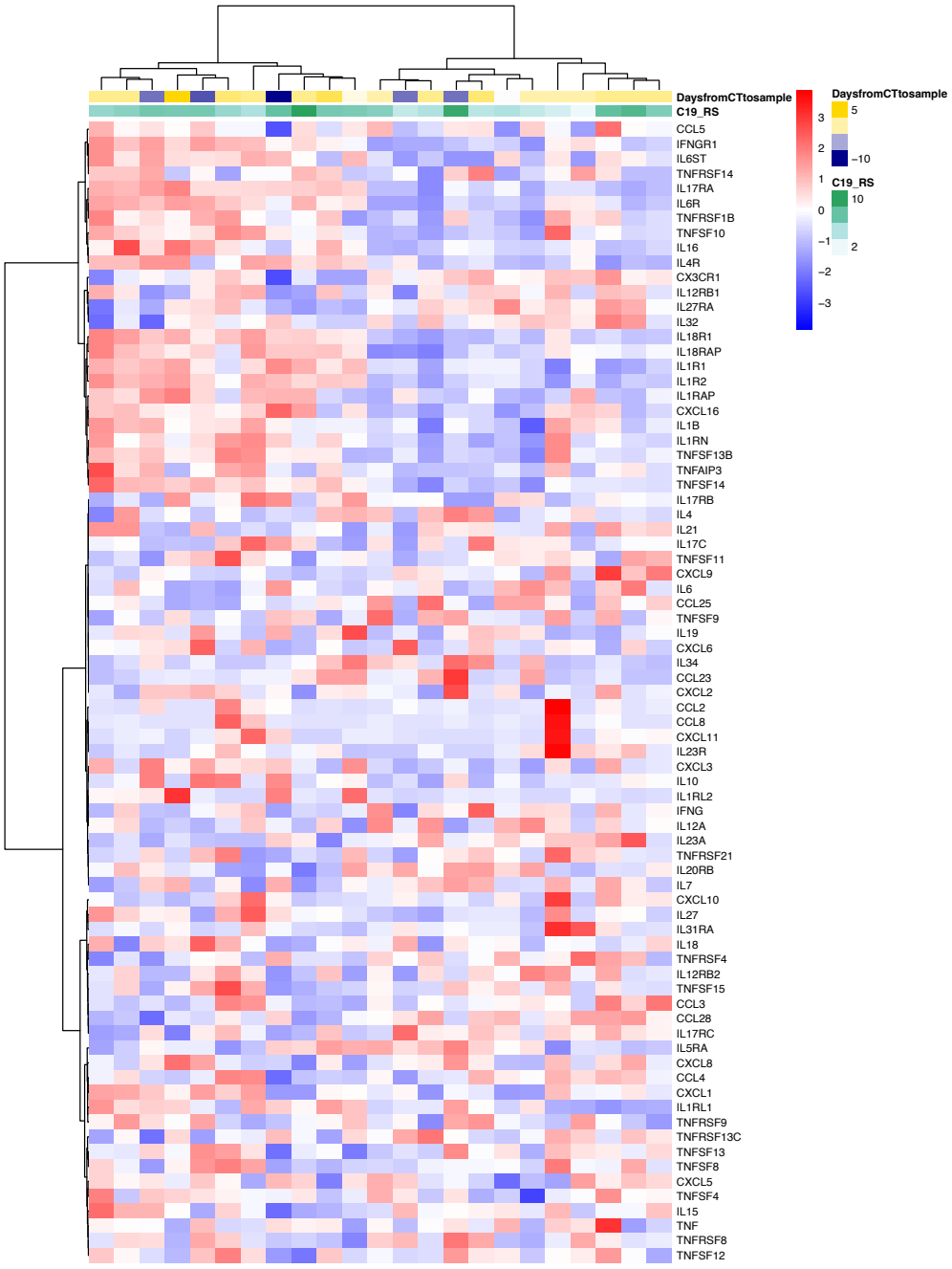
Online Figure 5 | Incremental value of C19-RS radiomic features in COVID-19 detection. Receiver operating characteristic curves of two logistic regression models showing that addition of the top third (n=8) of radiomic features comprising C19-RS significantly improved the performance of a baseline model consisting of age, sex, cardiovascular risk factors (hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease), C-reactive protein plasma levels, white blood cell count, plasma troponin, and history of chronic obstructive pulmonary disease for COVID-19 detection. P values derived from the DeLong test of areas under the curve for model 1 (age, sex, hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease, C-reactive protein plasma levels, white blood cell count, plasma troponin, and history of chronic obstructive pulmonary disease) and model 2 (model 1 plus eight C19-RS radiomic features with the highest gain values).



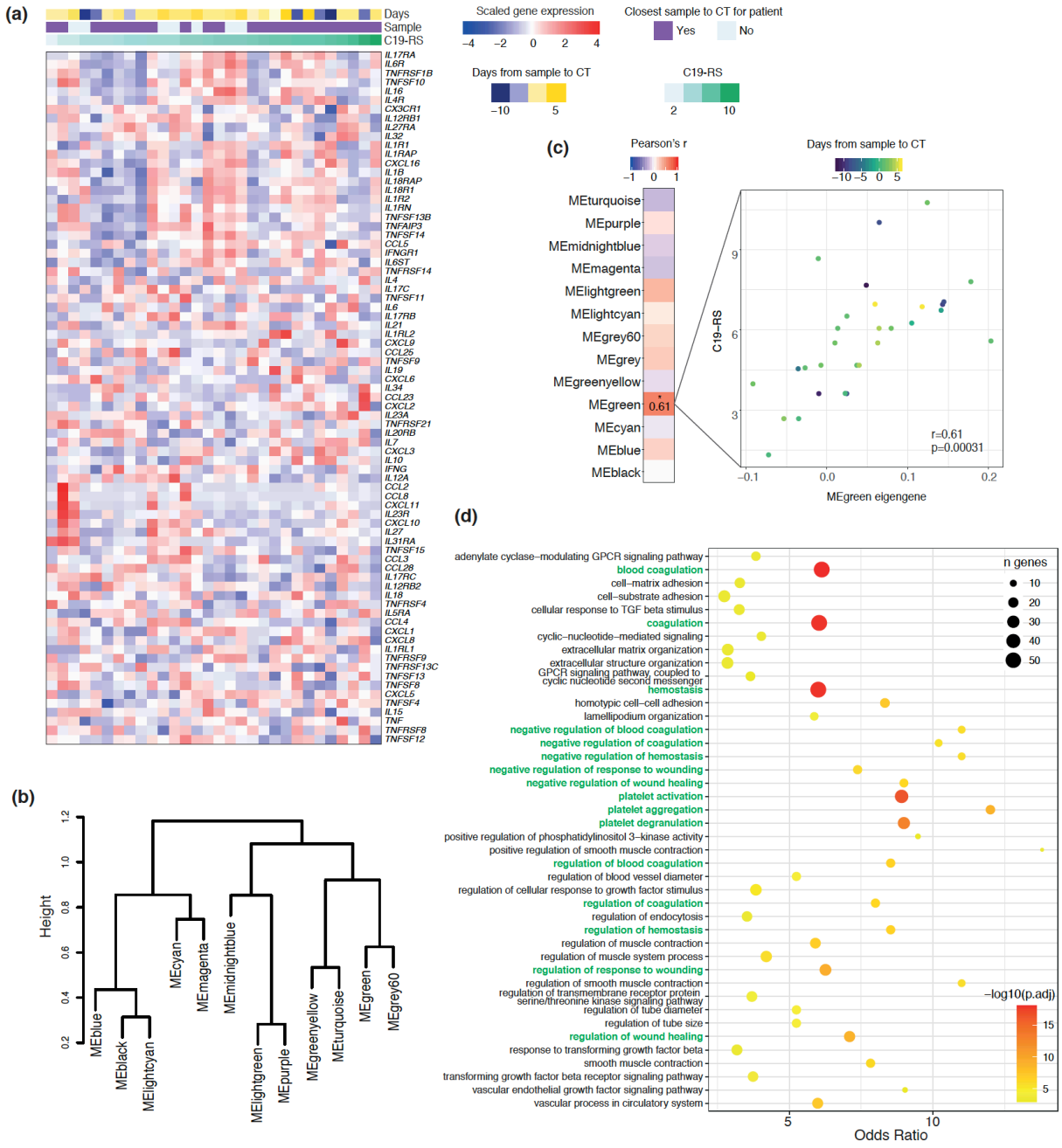
Online Figure 6 | Identifying the optimal C19-RS cut-off for in-hospital death prediction. Plot of the standardized log-rank statistic for prediction of death in-hospital versus different cut-off points for C19-RS, showing optimal discrimination for a cut-off point of 6.99, in the COVID-19 positive Study Arm 3 population (n=254).



Online Figure 7 | Decision curve analysis. Plot of decision curve analysis for composite endpoint prediction for each model. Model 1 (blue line) consists of demographic variables age, sex, hypertension (HTN), hyperlipidaemia (HLD), diabetes mellitus (DM), body mass index (BMI), presence of coronary artery disease (CAD), history of chronic obstructive pulmonary disease (COPD), tube voltage, and biochemistry biomarkers white blood cell count (WBC), C-reactive protein (CRP), and plasma troponin (Tn). Model 2 (purple line) includes all parameters in model 1 plus C19-RS. The y-axis measures net benefit, calculated by summing the benefits (true positives) and subtracting the harms (false positives), in which the latter are weighted by a factor related to the relative harm of a missed cancer compared with the harm of an unnecessary biopsy.

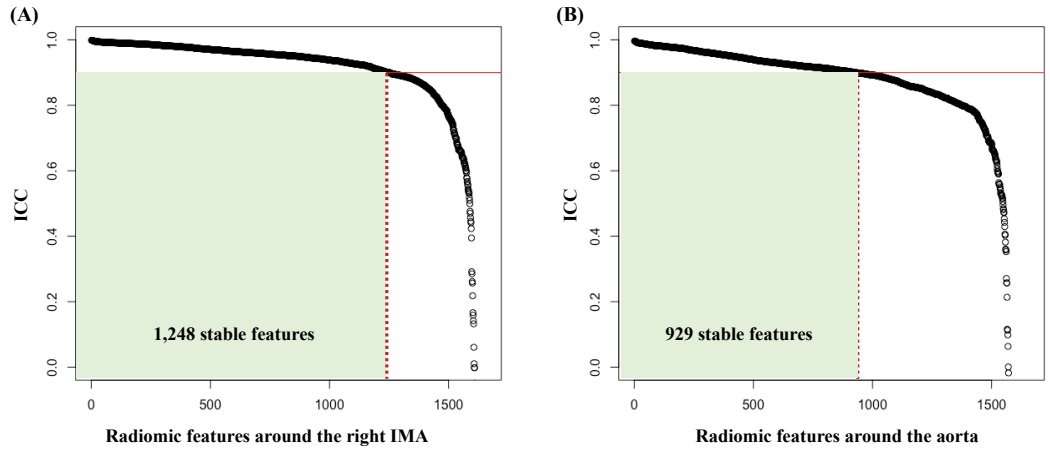


Online Figure 8 | Unsupervised hierarchical clustering of the inflammatory genes in whole blood in the COMBAT dataset. The scaled expression of the 77 genes retained after filtering is visualised for the closest sample per patient to the CT scan. Hierarchical clustering separates the patients into 2 clusters with no clear association with C19-RS.

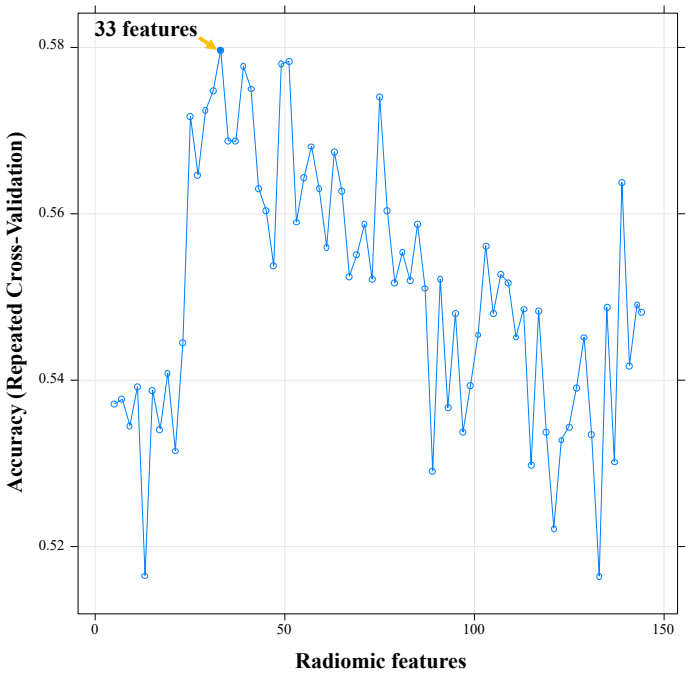


Online Figure 9 | Whole blood gene expression profiles and C19-RS. (a) Expression of the inflammation-related gene set in the COMBAT whole blood RNAseq dataset (30 samples from 23 COVID-19 patients from Study Arm 3 with CT scans, 77 genes detected). Samples are ordered by C19-RS, genes are clustered using Ward's method and Minkowski distance. Colour bar indicates sample timing and C19-RS. (b) Weighted Gene Correlation Network Analysis (WGCNA) on the

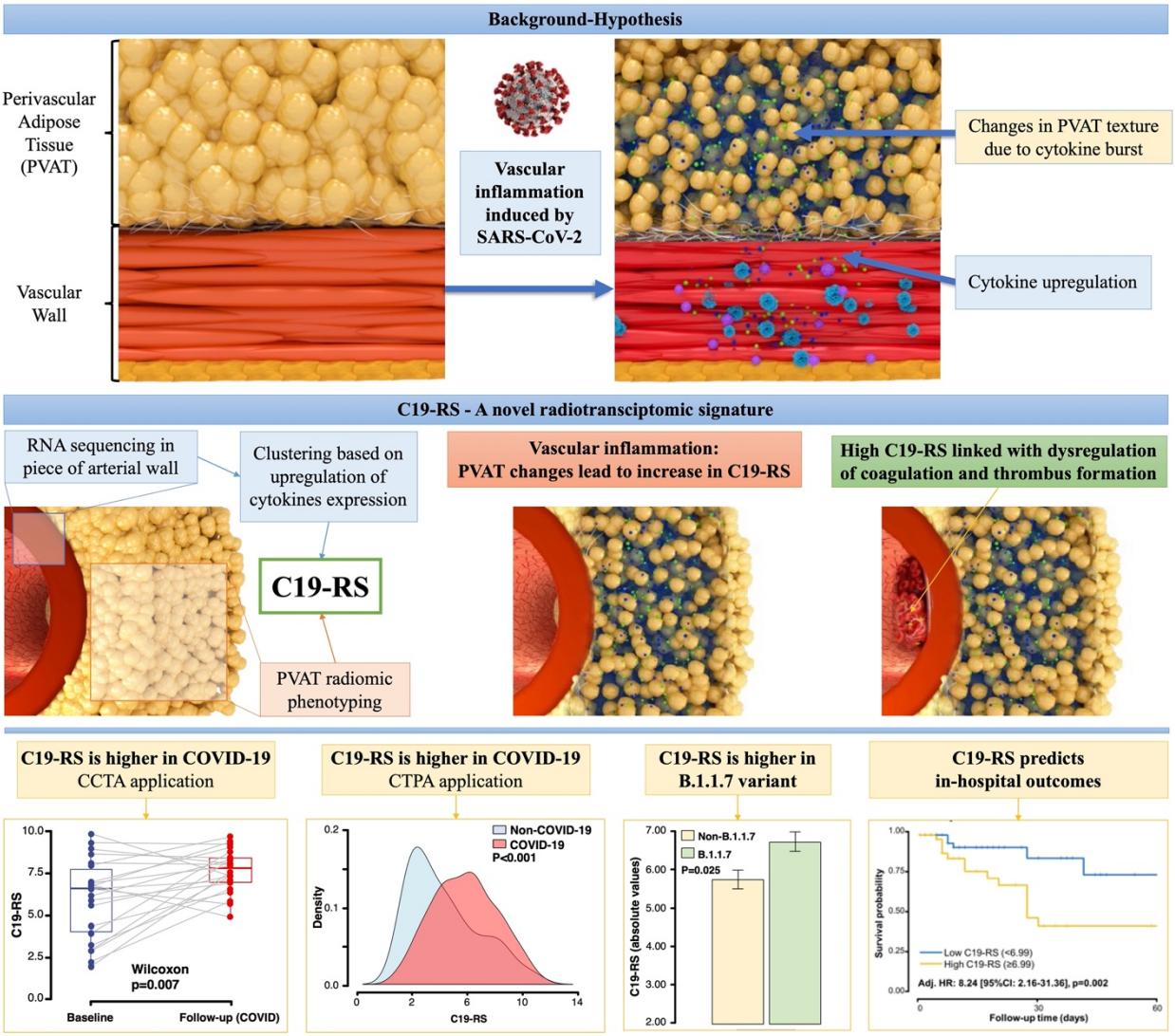
full COMBAT dataset defined gene expression modules2. Unassigned genes were categorised as MEgrey. (c) Correlation between C19-RS and the MEgreen eigengene by Pearson's r. (d) GOBP pathway enrichment analysis for MEgreen module member genes using hypergeometric test (terms with adjusted p-value <0.001 shown, full results in **appendix pp 65-67**). X axis (odds ratio) indicates strength of association between pathway and module memberships, size of point (n genes) indicates number of genes with overlapping membership, colour of points (adjusted p value) indicates significance of association.



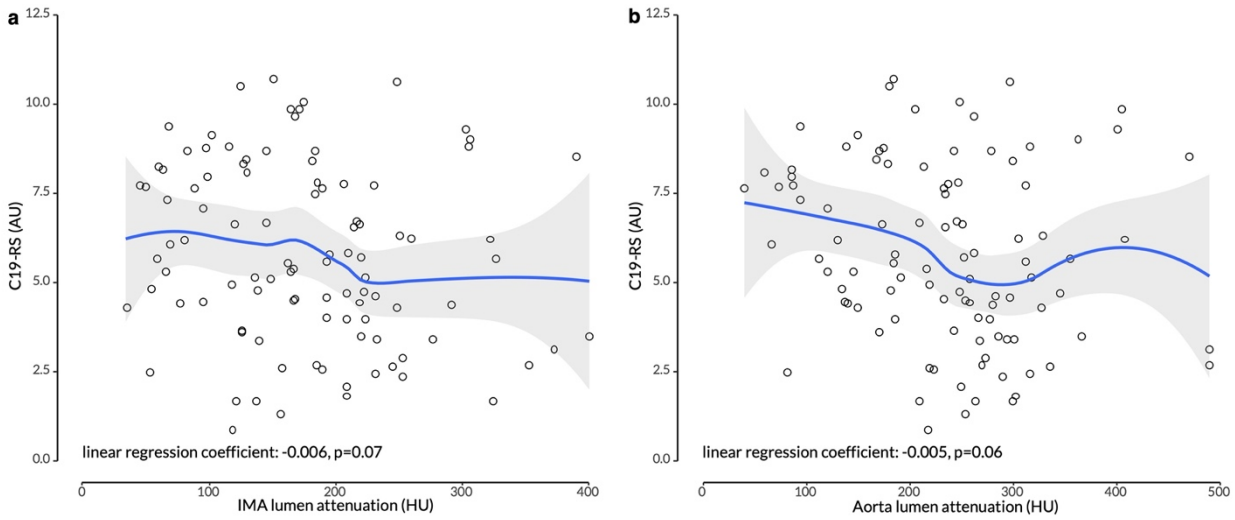
Online Figure 10 | Reproducibility analysis of perivascular adipose tissue features. Plot of the test-retest intraclass correlation coefficient (ICC) in the RIDER dataset of all 1,655 radiomic features measured around the right IMA (A) and descending aorta (B). Radiomic features are ranked on descending order based on their ICC value. A total of 1,248 and 929 radiomic features respectively were found to have an ICC equal to or greater than 0.90.



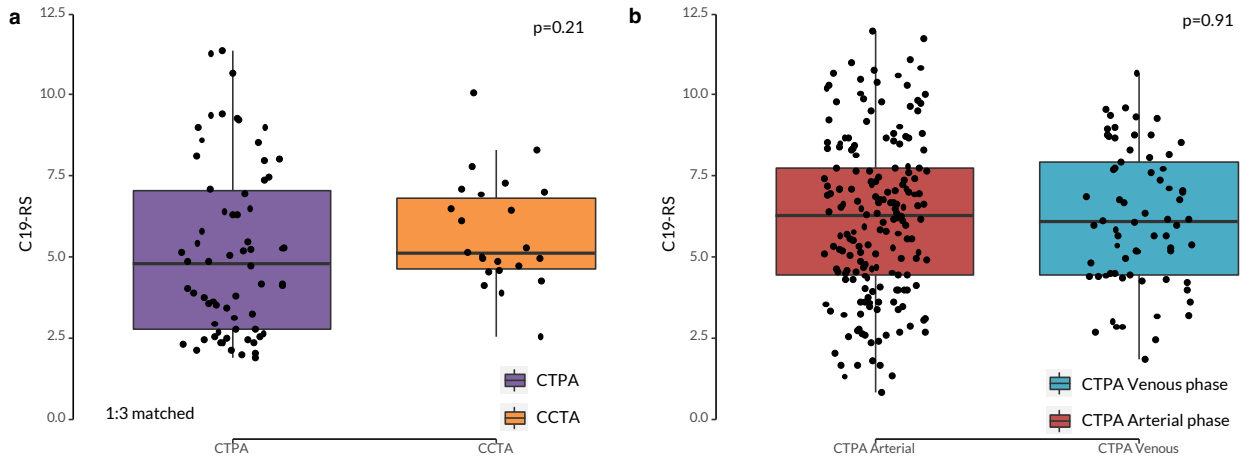
Online Figure 11 | Recursive feature elimination. Recursive feature elimination with a random forest algorithm and repeated five-fold cross-validation showed a plateau in the accuracy of the trained model with a maximal number of 33 selected features.



Online Figure 12 | The proposed imaging biomarker C19-RS. Vascular inflammation induced by the SARS-CoV-2 virus causes structural changes to perivascular adipose tissue. Utilising RNA sequencing data, a novel imaging biomarker -namely C19-RS- was trained to reflect upregulation of cytokine related genes in the arterial wall. C19-RS was higher in COVID-19 positive patients, particularly in those with the B.1.1.7 variant, and had significant prognostic value for in-hospital death prediction.



Online Figure 13 | Lumen attenuation and C19-RS. We tested lumen attenuation values in 100 consecutive patients from the Study Arm 3 population. C19-RS did not significantly correlate with either internal mammary artery (IMA) lumen attenuation (a) or thoracic aorta lumen attenuation (b) values. Linear regression models were fitted with C19-RS being the dependent variable and lumen attenuation values the independent variable for panels a and b respectively.



Online Figure 14 | C19-RS in CCTA vs CTPA and different CTPA phases. (a) We matched 22 COVID-19 negative patients undergoing coronary CTA from Study Arm 2 with 66 COVID-19 negative patients from Study Arm 3 undergoing pulmonary CTA, matched for age, sex, BMI, hypertension, and diabetes. No significant difference was observed in aboluste C19-RS values between the two imaging protocols. P value derived from the Mann-Whitney U test. (b) We further analysed quantitatively CTPAs in Study Arm 3 in order to stratify the cohort by CTPA phase (CTPA arterial phase: good visual contrast penetration vs CTPA venous phase: poor visual contrast penetration). No difference in C19-RS was observed. P value derived from the Mann-Whitney U test. We have further performed an additional sensitivity analysis in which we have included the phase as a co-variate in the prognostic modelling in Study Arm 3. Indeed, there was no impact of this parameter on the prognostic value of C19RS. (C19-RS Adj. HR 3.37 [95%CI: 1.64-6.93, p<0.001], adjusted for age above 65, sex, cardiovascular risk factors (hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease), C-reactive protein plasma levels, white blood cell count, plasma troponin, history of chronic obstructive pulmonary disease, CT tube voltage, and CTPA phase).

Online Table 1. List of genes chosen for unsupervised clustering in Study Arm 1

CCL1	CXCL13	IFNL4	IL1RL2	IL6R
CCL11	CXCL14	IFNW1	IL1RN	IL6ST
CCL13	CXCL16	IL10	IL2	IL7
CCL14	CXCL17	IL11	IL20	IL9
CCL15	CXCL2	IL12A	IL20RB	TNF
CCL16	CXCL3	IL12B	IL21	TNFAIP3
CCL17	CXCL5	IL12RB1	IL22	TNFRSF11B
CCL18	CXCL6	IL12RB2	IL23A	TNFRSF13C
CCL19	CXCL8	IL13	IL23R	TNFRSF14
CCL2	CXCL9	IL15	IL24	TNFRSF1B
CCL20	IFNA1	IL16	IL25	TNFRSF21
CCL21	IFNA10	IL17A	IL26	TNFRSF4
CCL22	IFNA14	IL17B	IL27	TNFRSF8
CCL23	IFNA16	IL17C	IL27RA	TNFRSF9
CCL24	IFNA17	IL17D	IL3	TNFSF10
CCL25	IFNA2	IL17F	IL31	TNFSF11
CCL26	IFNA21	IL17RA	IL31RA	TNFSF12
CCL27	IFNA4	IL17RB	IL32	TNFSF13
CCL28	IFNA5	IL17RC	IL33	TNFSF13B
CCL3	IFNA6	IL18	IL34	TNFSF14
CCL4	IFNA7	IL18R1	IL36A	TNFSF15
CCL5	IFNA8	IL18RAP	IL36B	TNFSF18
CCL7	IFNB1	IL19	IL36G	TNFSF4
CCL8	IFNE	IL1A	IL36RN	TNFSF8
CX3CL1	IFNG	IL1B	IL37	TNFSF9
CX3CR1	IFNGR1	IL1F10	IL4	
CXCL1	IFNK	IL1R1	IL4R	
CXCL10	IFNL1	IL1R2	IL5	
CXCL11	IFNL2	IL1RAP	IL5RA	
CXCL12	IFNL3	IL1RL1	IL6	

Online Table 2. Demographic characteristics of the Study Arm 1 population.

	Study Arm 1 population			
	All	**Cluster 1**	**Cluster 2**	**P value**
Total (n)	**55**	**28**	**27**	**-**
Age (years)	68.5[61.0-75.0]	68.0[58.8-75.0]	70.5[64.0-75.8]	0.467
Sex (male, %)	85.5	85.7	85.2	0.564
Risk factors (%)				
Hypertension	70.9	60.7	81.5	0.087
Dyslipidaemia	80.0	75.0	85.2	0.263
Diabetes mellitus	16.4	14.3	18.5	0.524
Smoking	56.4	57.2	55.5	0.928
Systolic blood pressure (mmHg)	130.02+/-17.92	129.19+/-19.48	130.92+/-16.42	0.731
Body Mass Index (BMI, kg/m^2)	27.82+/-4.21	27.14+/-4.63	28.52+/-3.68	0.236
Medication (%)				
Statins	87.3	85.7	88.9	0.913
ACEi	63.6	57.1	70.4	0.360
b-blockers	72.7	75.0	70.4	0.924
Nitrates	54.5	60.7	48.1	0.597
CT Scanning data				
Scan type	CCTA	CCTA	CCTA	
Tube Voltage (kVp) n (%)				
100	2	1	1	1
120	53	27	26	1

ACEi=angiotensin converting enzyme inhibitors; continuous variables reported as means+/-SEM or median [IQR], as appropriate. CCTA: Gated coronary computed tomography angiography.

Online Table 3. C19-RS RQS			
	Criteria	**Points**	**Comments**
1	Image protocol quality - well-documented image protocols (for example, contrast, slice thickness, energy, etc.) and/or usage of public image protocols allow reproducibility/replicability	1/2	All image protocols documented clearly on Online Appendix
2	Multiple segmentations - possible actions are: segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyse feature robustness to segmentation variabilities	1/1	Three-dimensional segmentation performed by two independent analysts. Scan-rescan stability assessment of all 3,310 radiomic features done against the RIDER dataset.
3	Phantom study on all scanners - detect inter-scanner differences and vendor-dependent features. Analyse feature robustness to these sources of variability	0/1	No phantom studies performed.
4	Imaging at multiple time points - collect images of individuals at additional time points. Analyse feature robustness to temporal variabilities (for example, organ movement, organ expansion/shrinkage)	1/1	Imaging at multiple time points given in Figure 4.
5	Feature reduction or adjustment for multiple testing - decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consider feature robustness when selecting features	3/3	Feature reduction performed against scan-rescan stability, and interobserver consistency. In addition, radiomic features that were significantly correlated with BMI and total intrathoracic adipose tissue were removed.
6	Multivariable analysis with non radiomics features (for example, EGFR mutation) - is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non radiomics features	1/1	All models adjusted for age, sex, cardiovascular risk factors (hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease), C-reactive protein plasma levels, white blood cell count, plasma troponin, history of chronic obstructive pulmonary disease, and tube voltage.
7	Detect and discuss biological correlates - demonstration of phenotypic differences (possibly associated with underlying gene–protein expression patterns) deepens understanding of radiomics and biology	1/1	Biological meaning of radiomic features within C19-RS given in Figure 4a.
8	Cut-off analyses - determine risk groups by either the median, a previously published cut-off or report a continuous risk variable. Reduces the risk of reporting overly optimistic results	1/1	C19-RS cut-off of 6.99 identified by the value that maximized the log-rank statistic for death in hospital.

9	Discrimination statistics - report discrimination statistics (for example, C-statistic, ROC curve, AUC) and their statistical significance (for example, p-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	2/2	Discrimination statistics and ROC curves for C19-RS presented in Figure 5a.
10	Calibration statistics - report calibration statistics (for example, Calibration-in-the-large/slope, calibration plots) and their statistical significance (for example, *P*-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	1/2	C19-RS feature selection performed by recursive feature elimination with a random forest algorithm and repeated five-fold cross-validation.
11	Prospective study registered in a trial database - provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker	7/7	Ongoing C19-RS validation/testing in the RECOVERY trial (NCT04381936).
12	Validation - the validation is performed without retraining and without adaptation of the cut-off value, provides crucial information with regard to credible clinical performance	5/5	External validation based on three datasets from distinct institutes (Oxford, Bath, Leicester).
13	Comparison to 'gold standard' - assess the extent to which the model agrees with/is superior to the current 'gold standard' method (for example, TNM-staging for survival prediction). This comparison shows the added value of radiomics	2/2	Presented in Figure 5b.
14	Potential clinical utility - report on the current and potential application of the model in a clinical setting (for example, decision curve analysis).	2/2	A decision curve analysis is presented in Online Figure 6.
15	Cost-effectiveness analysis - report on the cost-effectiveness of the clinical application (for example, QALYs generated)	0/1	No within the scope of this study.
16	Open science and data - make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study	0/4	Individual participant-level data used for this report are not publicly available, because they contain protected patient health information. Requests for data access should be directed to the corresponding author via email.
	Total points (28/36 = 77.8%)		

Online Table 4. Demographic characteristics of the Study Arm 2 population.

	Study Arm 2 population			
	All	**COVID**	**Controls**	**P value**
Total (n)	**44**	**22**	**22**	**-**
Age (years)	57.5 [48.0, 63.3]	58.0 [48.8, 66.0]	56.5 [48.3, 62.8]	0.518
Sex (male, %)	72.7	72.7	72.7	1
Risk factors (%)				
Hypertension	50.0	50.0	50.0	1
Diabetes mellitus	25.0	18.2	31.8	0.486
Body Mass Index (kg/m^2)	32.4+/-6.3	32.5+/-7.7	32.3+/-4.7	0.911
COVID severity				
Mild	-	40.9%	-	-
Moderate	-	36.4%	-	-
Severe	-	13.6%	-	-
Critical	-	9.1%	-	-
C19-RS Baseline	5.9 [4.5, 6.9]	6.5 [3.9, 7.6]	5.1 [4.6, 6.8]	0.453
C19-RS Follow-up	6.6 [4.7, 8.0]	7.7 [6.9, 8.3]	4.6 [3.7, 6.6]	<0.001
Delta C19-RS	-	1.7 [-0.1, 3.2]	-0.3 [-1.8, 0.9]	0.005
Time between scans (years)	-	2.2 [1.8, 4.1]	4.3 [4.1, 4.4]	0.002

Continuous variables reported as means+/-SEM or median [IQR], as appropriate. Factor variables are presented as percentages. Delta C19-RS describes the difference in C19-RS between baseline and follow-up scanning.

Online Table 5. Demographic characteristics of the Study Arm 3 population.

	First UK Wave					Second UK Wave
	Strata by random split		Strata by COVID status			All
	Exploratory (20%)	Validation (80%)	COVID	Non-COVID	p value	Validation (100%)
	N=53	N=216	N=169	N=100		N=115
Age (years)	63.4[50.7-76.0]	57.9[46.6-74.0]	60.0 [51.1-75.7]	53.2 [40.1-69.8]	0.001	62.7 [51.6-75.6]
Sex (male, %)	26 (49.1)	95 (44.0)	96 (56.8)	42 (46.7)	0.154	72 (62.6)
Risk factors, n (%)						
Hypertension	23 (43.4)	69 (31.9)	72 (43.4)	20 (25.3)	0.01	26 (22.6)
Dyslipidaemia	6 (11.3)	31 (14.4)	25 (15.2)	12 (15.2)	1	16 (13.9)
Diabetes mellitus	13 (24.5)	37 (17.1)	39 (23.5)	11 (13.9)	0.117	25 (21.7)
COPD	2 (3.8)	12 (5.6)	10 (6.1)	4 (5.1)	0.985	9 (7.8)
CAD	5 (9.4)	16 (7.4)	13 (7.9)	8 (10.3)	0.71	15 (13.0)
SBP (mmHg)	130.9±27.3	132.5±22.0	129.7±21.4	137.1±257	0.021	129.7±18.6
BMI (kg/m^2)	30.0±9.6	27.7±6.3	28.5±6.1	27.2±9.4	0.298	30.6±7.6
Medication, n (%)						
Statins	13 (24.5)	59 (27.3)	55 (33.1)	17 (21.5)	0.086	18 (15.7)
ACEi	7 (13.2)	27 (12.5)	29 (17.5)	5 (6.3)	0.031	10 (8.7)
Beta blockers	5 (9.4)	23 (10.6)	20 (12.0)	8 (10.1)	0.82	9 (7.8)
Biochemical measurements						
CRP (mg/L)	82.9[33.0-180.0]	75.6[18.0-156.3]	97.0 [37.5, 185.3]	16.70 [2.1, 86.5]	<0.001	98.8 [54.9, 141.4]
WBC ($x10^3$)	8.3 [6.4-12.0]	8.6 [6.1-12.3]	7.5 [5.4, 10.7]	10.3 [7.5, 14.7]	<0.001	7.1 [5.3, 9.5]
Troponin (ng/L)	5.0 [2.0-62.5]	5.0 [2.0-20.0]	5.0 [2.0, 32.0]	3.0 [2.0, 11.0]	0.255	6.0 [4.0, 17.5]
COVID Severity						
Mild	-	-	34.3%	-	-	9.6%
Moderate	-	-	36.1%	-	-	49.6%
Severe	-	-	10.0%	-	-	25.2%
Critical	-	-	19.5%	-	-	15.6%
Scan parameters						
Tube Voltage (kVp) or Effective Energy (keV), n (%)						
100 kVp	35 (66.0)	142 (65.7)	114 (67.5)	63 (63.0)	0.45	98 (85.2)
110 kVp	0 (0)	0 (0)	0	0	-	13 (11.3)
120 kVp	8 (15.1)	27 (12.5)	1 (0.6)	34 (34.0)	<0.001	4 (3.5)
55 keV	1 (1.9)	3 (1.4)	3 (1.8)	1 (1.0)	0.61	0 (0)
58 keV	9 (17.0)	40 (18.5)	48 (28.4)	1 (1.0)	<0.001	0 (0)
70 keV	0 (0.0)	4 (1.9)	3 (1.8)	1 (1.0)	0.61	0 (0)
Outcomes						
Days in hospital	7.0[0.0-18.0]	4.0[0.0-13.0]	8.0 [3.0, 18.0]	0.0 [0.0, 4.0]	<0.001	8.0 [5.0, 14.0]
Death in-hospital, n(%)	12 (22.6)	21 (9.7)	30 (17.8)	3 (3.3)	0.002	19 (16.5)
Composite endpoint, n(%)	19 (35.8)	59 (27.3)	70 (41.4)	8 (8.9)	<0.001	30 (26.1)

COPD=Chronic Obstructive pulmonary disease; CAD: Coronary Artery Disease; SBP: Systolic blood pressure; ACEi=angiotensin converting enzyme inhibitors; BMI=Body Mass Index; WBC=White Blood Cells count; ICU=Intensive care unit; Composite endpoint includes death in-hospital and/or admission to intensive care unit; continuous variables reported as median [IQR]. Continuous variables are expressed as mean±SD or median[range] as appropriate. P values derived from comparisons between COVID and non-COVID patients of the first wave Study Arm 3 population.

Online Table 6. Demographic characteristics of the Study Arm 4 population.

	Study Arm 4 population
Total (n)	**104**
Recruitment centre/region	
Leicester	56
Bath	48
Age (years)	63.7 [54.0, 74.0]
Male sex, n (%)	63 (60.6)
Risk factors, n (%)	
Hypertension	30 (28.8)
Dyslipidaemia	13 (12.5)
Diabetes mellitus	28 (26.9)
CAD	3 (2.9)
COPD	17 (16.3)
Systolic blood pressure (mmHg)	128.2 ± 22.4
Body Mass Index (kg/m^2)	29.7 ± 5.8
Biochemical measurements	
C-reactive protein (mg/L)	62.5 [28.0, 148.5]
Troponin	9.5 [6.0, 23.3]
White blood cell count ($x10^3$)	9.4 [6.5, 12.0]
Lymphocytes cell count ($x10^3$)	1.0 [0.7, 1.6]
Monocytes cell count ($x10^3$)	0.5 [0.3, 0.8]
Tube Voltage (kVp)	
100 (kVp, n, %)	55 (52.9)
110 (kVp, n, %)	3 (2.9)
120 (kVp, n, %)	39 (37.5)
140 (kVp, n, %)	7 (6.7)
Severity	
Mild	13.5%
Moderate	39.4%
Severe	27.9%
Critical	19.2%

COPD=Chronic Obstructive pulmonary disease; CAD: Coronary Artery Disease; continuous variables reported as means+/-SD or median[range] as appropriate.

Online Table 7. Conversion factors for different CT energies

	Conversion Factor for Adipose Tissue
100 kVp (tube voltage)	1 (reference)
110 kVp (tube voltage)	1.054740019
120 kVp (tube voltage)	1.114849188
70keV (effective energy)	0.988683128
58 keV (effective energy)	0.817537672
55 keV (effective energy)	0.771195642

Online Table 8. Comparison of performance of different machine learning approached for the development of C19-RS.

Model-Method	AUC for COVID-19 status detection in Study Arm 3 (n=331)
Extreme Gradient Boosting	
gbtree	0.65 [95%CI: 0.57-0.73], p<0.001
gblinear	0.57 [95%CI: 0.50-0.64], p=0.056
dart	0.62 [95%CI: 0.55-0.70], p<0.001
Random Forest	
rf	0.62 [95%CI: 0.54-0.70], p<0.001
ranger	0.58 [95%CI: 0.51-0.66], p=0.02
Neural Network	
avNNet	0.55 [95%CI: 0.47-0.62], p=0.18
Bayesian Model	
bayesglm	0.56 [95%CI: 0.49-0.63], p=0.09

Online Table 9. Radiomic features comprising C19-RS

AA	Radiomic Feature Name
F1	wavelet.LHH_glcm_Imc2_aorta
F2	log.sigma.4.0.mm.3D_glszm_GrayLevelNonUniformityNormalized_ima
F3	wavelet.LHH_glcm_Correlation_aorta
F4	wavelet.LLL_glcm_JointAverage_ima
F5	log.sigma.1.0.mm.3D_ngtdm_Strength_aorta
F6	wavelet.LLL_glcm_DifferenceVariance_aorta
F7	log.sigma.2.0.mm.3D_glcm_Correlation_aorta
F8	log.sigma.1.0.mm.3D_firstorder_Skewness_aorta
F9	log.sigma.1.0.mm.3D_glszm_SizeZoneNonUniformity_aorta
F10	logarithm_glcm_Imc2_ima
F11	wavelet.LLL_glszm_GrayLevelNonUniformityNormalized_ima
F12	log.sigma.5.0.mm.3D_glrlm_RunLengthNonUniformityNormalized_aorta
F13	wavelet.HLH_firstorder_Maximum_ima
F14	log.sigma.5.0.mm.3D_glszm_SmallAreaHighGrayLevelEmphasis_aorta
F15	wavelet.LHL_firstorder_Median_aorta
F16	log.sigma.3.0.mm.3D_firstorder_Maximum_ima
F17	log.sigma.1.0.mm.3D_glcm_InverseVariance_ima
F18	wavelet.LHL_firstorder_Range_aorta
F19	log.sigma.4.0.mm.3D_firstorder_Median_aorta
F20	wavelet.HHH_glcm_ClusterTendency_ima
F21	wavelet.HLL_glcm_ClusterProminence_aorta
F22	wavelet.HHH_glcm_MaximumProbability_aorta
F23	wavelet.LLL_glrlm_RunEntropy_aorta
F24	wavelet.HLL_ngtdm_Strength_aorta
F25	wavelet.HHL_firstorder_Maximum_aorta

Online Table 10. Proportion of Study Arm 3 patients with high C19-RS and outcomes with low troponin and CRP.

	Total number	Deaths in hospital	Composite endpoint	High C19-RS
CRP<50mg/L	73	9 (12.3%)	14 (19.1%)	18 (24.6%) 5 deaths/6 composite
Troponin<20ng/L	182	16 (8.7%)	48 (26.37%)	61 (33.5%) 11 deaths/23 composite

CRP: C-Reactive Protein.

Online Table 11. Sensitivity analysis excluding variables with high missingness

	Study Arm 3	Study Arm 4
Fully Adjusted HR*	3.31 [95%CI: 1.49-7.33], p=0.003	2.58 [95%CI: 1.10-6.05], p=0.028
Sensitivity Analysis HR#	3.55 [95%CI: 1.73-7.29], p=0.0006	2.21 [95%CI: 1.04-4.71], p=0.04

Sensitivity analyses in Study Arm 3 and 4 populations are exploratory post-hoc analyses excluding variables with high degree of missingness (>10%) after data collection was completed.

*adjusted for adjusted for age above 65, sex, hypertension, hyperlipidaemia, diabetes, BMI, presence of coronary artery disease, C-reactive protein plasma levels, white blood cell count, plasma troponin, history of chronic obstructive pulmonary disease and CT tube voltage

#adjusted only for variables with missingness below 10%: age above 65, sex, hypertension, hyperlipidaemia, diabetes, presence of coronary artery disease, white blood cell count, history of chronic obstructive pulmonary disease and CT tube voltage

Online Table 12. Genes in the Green Module

ENSEMBL ID	ENTREZ ID
ENSG00000148498	PARD3
ENSG00000088726	TMEM40
ENSG00000047648	ARHGAP6
ENSG00000101162	TUBB1
ENSG00000005249	PRKAR2B
ENSG00000085733	CTTN
ENSG00000127533	F2RL3
ENSG00000061918	GUCY1B1
ENSG00000151693	ASAP2
ENSG00000107863	ARHGAP21
ENSG00000065534	MYLK
ENSG00000177119	ANO6
ENSG00000154146	NRGN
ENSG00000123739	PLA2G12A
ENSG00000138798	EGF
ENSG00000144677	CTDSPL
ENSG00000173210	ABLIM3
ENSG00000140022	STON2
ENSG00000011105	TSPAN9
ENSG00000204424	LY6G6F
ENSG00000102230	PCYT1B
ENSG00000138735	PDE5A
ENSG00000171611	PTCRA
ENSG00000088053	GP6
ENSG00000013016	EHD3
ENSG00000107438	PDLIM1
ENSG00000101856	PGRMC1
ENSG00000178033	CALHM5
ENSG00000197461	PDGFA
ENSG00000158457	TSPAN33
ENSG00000049323	LTBP1
ENSG00000168497	CAVIN2
ENSG00000090975	PITPNM2
ENSG00000095303	PTGS1
ENSG00000187800	PEAR1
ENSG00000102362	SYTL4

ENSG00000164116	GUCY1A1
ENSG00000169756	LIMS1
ENSG00000108839	ALOX12
ENSG00000180354	MTURN
ENSG00000072422	RHOBTB1
ENSG00000166091	CMTM5
ENSG00000183785	TUBA8
ENSG00000225528	Z82206.1
ENSG00000204420	MPIG6B
ENSG00000128266	GNAZ
ENSG00000160789	LMNA
ENSG00000185245	GP1BA
ENSG00000113140	SPARC
ENSG00000205038	PKHD1L1
ENSG00000198478	SH3BGRL2
ENSG00000174175	SELP
ENSG00000149218	ENDOD1
ENSG00000081377	CDC14B
ENSG00000149564	ESAM
ENSG00000165914	TTC7B
ENSG00000165702	GFI1B
ENSG00000164181	ELOVL7
ENSG00000111644	ACRBP
ENSG00000152952	PLOD2
ENSG00000035403	VCL
ENSG00000169247	SH3TC2
ENSG00000267279	AC090409.1
ENSG00000061676	NCKAP1
ENSG00000169704	GP9
ENSG00000099256	PRTFDC1
ENSG00000179855	GIPC3
ENSG00000006638	TBXA2R
ENSG00000166681	BEX3
ENSG00000188191	PRKAR1B
ENSG00000143363	PRUNE1
ENSG00000146376	ARHGAP18
ENSG00000185532	PRKG1
ENSG00000197959	DNM3
ENSG00000165646	SLC18A2
ENSG00000259207	ITGB3

ENSG00000180190	TDRP
ENSG00000185340	GAS2L1
ENSG00000163898	LIPH
ENSG00000187699	C2orf88
ENSG00000269556	TMEM185A
ENSG00000172572	PDE3A
ENSG00000128245	YWHAH
ENSG00000122786	CALD1
ENSG00000204310	AGPAT1
ENSG00000109265	CRACD
ENSG00000100351	GRAP2
ENSG00000108960	MMD
ENSG00000259330	INAFM2
ENSG00000108576	SLC6A4
ENSG00000161911	TREML1
ENSG00000182732	RGS6
ENSG00000160014	CALM3
ENSG00000124491	F13A1
ENSG00000212864	RNF208
ENSG00000100678	SLC8A3
ENSG00000177076	ACER2
ENSG00000074416	MGLL
ENSG00000069966	GNB5
ENSG00000246889	AP000487.1
ENSG00000163737	PF4
ENSG00000127920	GNG11
ENSG00000166963	MAP1A
ENSG00000254703	SENCR
ENSG00000106665	CLIP2
ENSG00000150630	VEGFC
ENSG00000119280	C1orf198
ENSG00000108846	ABCC3
ENSG00000260244	AC104083.1
ENSG00000259719	LINC02284
ENSG00000147036	LANCL3
ENSG00000103184	SEC14L5
ENSG00000137672	TRPC6
ENSG00000138722	MMRN1
ENSG00000104341	LAPTM4B
ENSG00000169398	PTK2

ENSG00000137941	TTLL7
ENSG00000279970	AC023024.2
ENSG00000111328	CDK2AP1
ENSG00000003436	TFPI
ENSG00000128578	STRIP2
ENSG00000176783	RUFY1
ENSG00000248516	AC105415.1
ENSG00000197415	VEPH1
ENSG00000120885	CLU
ENSG00000082781	ITGB5
ENSG00000110013	SIAE
ENSG00000140374	ETFA
ENSG00000102178	UBL4A
ENSG00000187098	MITF
ENSG00000133627	ACTR3B
ENSG00000140479	PCSK6
ENSG00000127831	VIL1
ENSG00000250091	DNAH10OS
ENSG00000148426	PROSER2
ENSG00000088826	SMOX
ENSG00000162367	TAL1
ENSG00000255240	AP001636.3
ENSG00000150594	ADRA2A
ENSG00000069535	MAOB
ENSG00000081181	ARG2
ENSG00000205309	NT5M
ENSG00000133401	PDZD2
ENSG00000188677	PARVB
ENSG00000106976	DNM1
ENSG00000198873	GRK5
ENSG00000101082	SLA2
ENSG00000103942	HOMER2
ENSG00000047617	ANO2
ENSG00000160013	PTGIR
ENSG00000225936	AL731557.1
ENSG00000140682	TGFB1I1
ENSG00000005961	ITGA2B
ENSG00000178057	NDUFAF3
ENSG00000256235	SMIM3
ENSG00000101335	MYL9

ENSG00000259869	AL022344.1
ENSG00000137942	FNBP1L
ENSG00000128791	TWSG1
ENSG00000227811	INKA2-AS1
ENSG00000163430	FSTL1
ENSG00000176903	PNMA1
ENSG00000102804	TSC22D1
ENSG00000138685	FGF2
ENSG00000269970	AL162424.1
ENSG00000022267	FHL1
ENSG00000153162	BMP6
ENSG00000167553	TUBA1C
ENSG00000167645	YIF1B
ENSG00000124762	CDKN1A
ENSG00000235257	ITGA9-AS1
ENSG00000264964	AP001033.3
ENSG00000254614	AP003068.2
ENSG00000234810	AL603840.1
ENSG00000253520	AC136628.3
ENSG00000153071	DAB2
ENSG00000203485	INF2
ENSG00000188549	CCDC9B
ENSG00000110880	CORO1C
ENSG00000117155	SSX2IP
ENSG00000231131	LNCAROD
ENSG00000095321	CRAT
ENSG00000198948	MFAP3L
ENSG00000050393	MCUR1
ENSG00000249898	MCPH1-AS1
ENSG00000198513	ATL1
ENSG00000166086	JAM3
ENSG00000167641	PPP1R14A
ENSG00000119862	LGALSL
ENSG00000188013	MEIS3P2
ENSG00000106366	SERPINE1
ENSG00000178732	GP5
ENSG00000004866	ST7
ENSG00000132970	WASF3
ENSG00000147650	LRP12
ENSG00000182253	SYNM

ENSG00000100439	ABHD4
ENSG00000276141	WHAMMP3
ENSG00000109572	CLCN3
ENSG00000165682	CLEC1B
ENSG00000254877	AP001636.2
ENSG00000150637	CD226
ENSG00000071051	NCK2
ENSG00000041353	RAB27B
ENSG00000253819	LINC01151
ENSG00000174456	C12orf76
ENSG00000134548	SPX
ENSG00000143369	ECM1
ENSG00000156011	PSD3
ENSG00000143595	AQP10
ENSG00000129521	EGLN3
ENSG00000261253	AC137932.2
ENSG00000260912	AL158206.1
ENSG00000197879	MYO1C
ENSG00000146858	ZC3HAV1L
ENSG00000198586	TLK1
ENSG00000185614	INKA1
ENSG00000198753	PLXNB3
ENSG00000140416	TPM1
ENSG00000142192	APP
ENSG00000141198	TOM1L1
ENSG00000127838	PNKD
ENSG00000169083	AR
ENSG00000173542	MOB1B
ENSG00000253394	LINC00534
ENSG00000231652	AL590428.1
ENSG00000197122	SRC
ENSG00000117400	MPL
ENSG00000154917	RAB6B
ENSG00000205126	ACCSL
ENSG00000184602	SNN
ENSG00000097021	ACOT7
ENSG00000214018	RRM2P3
ENSG00000169047	IRS1
ENSG00000184500	PROS1
ENSG00000172794	RAB37

ENSG00000223855	AC147651.1
ENSG00000112078	KCTD20
ENSG00000235609	AF127577.4
ENSG00000141873	SLC39A3
ENSG00000185052	SLC24A3
ENSG00000024422	EHD2
ENSG00000158560	DYNC1I1
ENSG00000233276	GPX1
ENSG00000079482	OPHN1
ENSG00000183914	DNAH2
ENSG00000172889	EGFL7
ENSG00000260792	LINC02280
ENSG00000171159	C9orf16
ENSG00000068796	KIF2A
ENSG00000109089	CDR2L
ENSG00000084693	AGBL5
ENSG00000283633	AP000547.3
ENSG00000248636	AC002070.1
ENSG00000175854	SWI5
ENSG00000133317	LGALS12
ENSG00000185015	CA13
ENSG00000105971	CAV2
ENSG00000185222	TCEAL9
ENSG00000156535	CD109
ENSG00000103876	FAH
ENSG00000126903	SLC10A3
ENSG00000179399	GPC5
ENSG00000213977	TAX1BP3
ENSG00000176490	DIRAS1
ENSG00000139835	GRTP1
ENSG00000101412	E2F1
ENSG00000235513	L3MBTL2-AS1
ENSG00000167363	FN3K
ENSG00000272468	AL021807.1
ENSG00000072657	TRHDE
ENSG00000267060	PTGES3L
ENSG00000186889	TMEM17
ENSG00000134909	ARHGAP32
ENSG00000170935	NCBP2L
ENSG00000255478	AP000944.1

ENSG00000125257	ABCC4
ENSG00000112290	WASF1
ENSG00000167460	TPM4
ENSG00000169241	SLC50A1
ENSG00000198829	SUCNR1
ENSG00000173852	DPY19L1
ENSG00000255045	AP000866.5
ENSG00000225981	AC102953.1
ENSG00000213672	NCKIPSD
ENSG00000148484	RSU1
ENSG00000131711	MAP1B
ENSG00000127325	BEST3
ENSG00000188580	NKAIN2
ENSG00000273123	AC020634.2
ENSG00000224805	LINC00853
ENSG00000112977	DAP
ENSG00000143195	ILDR2
ENSG00000135919	SERPINE2
ENSG00000115290	GRB14
ENSG00000108039	XPNPEP1
ENSG00000250334	LINC00989
ENSG00000162687	KCNT2
ENSG00000172159	FRMD3
ENSG00000147394	ZNF185
ENSG00000006831	ADIPOR2
ENSG00000143418	CERS2
ENSG00000101333	PLCB4
ENSG00000173626	TRAPPC3L
ENSG00000163736	PPBP
ENSG00000103740	ACSBG1
ENSG00000169946	ZFPM2
ENSG00000134668	SPOCD1
ENSG00000120279	MYCT1
ENSG00000284693	LINC02606
ENSG00000282863	AC012560.2
ENSG00000162852	CNST
ENSG00000017260	ATP2C1
ENSG00000228474	OST4
ENSG00000087460	GNAS
ENSG00000244041	LINC01011

ENSG00000153714	LURAP1L
ENSG00000138080	EMILIN1
ENSG00000144893	MED12L
ENSG00000214357	NEURL1B
ENSG00000156206	CFAP161
ENSG00000072042	RDH11
ENSG00000235927	NEXN-AS1
ENSG00000103316	CRYM
ENSG00000253227	AC090192.2
ENSG00000183671	GPR1
ENSG00000169860	P2RY1
ENSG00000115170	ACVR1
ENSG00000165716	DIPK1B
ENSG00000174099	MSRB3
ENSG00000113361	CDH6
ENSG00000255874	LINC00346
ENSG00000167468	GPX4
ENSG00000130958	SLC35D2
ENSG00000109066	TMEM104
ENSG00000265148	TSPOAP1-AS1
ENSG00000273117	AC144652.1
ENSG00000180914	OXTR
ENSG00000126803	HSPA2
ENSG00000143995	MEIS1
ENSG00000146416	AIG1
ENSG00000100994	PYGB
ENSG00000127252	PLAAT1
ENSG00000169313	P2RY12
ENSG00000287815	AC005081.1
ENSG00000127526	SLC35E1
ENSG00000181104	F2R
ENSG00000226510	UPK1A-AS1
ENSG00000182048	TRPC2
ENSG00000115641	FHL2
ENSG00000157837	SPPL3
ENSG00000054356	PTPRN
ENSG00000237118	CYP2F2P
ENSG00000240497	AC092919.2
ENSG00000250641	LY6G6F-LY6G6D

ENSG00000167100	SAMD14
ENSG00000196924	FLNA
ENSG00000254786	AP001636.1
ENSG00000183963	SMTN
ENSG00000180573	H2AC6
ENSG00000249992	TMEM158
ENSG00000206052	DOK6
ENSG00000180694	TMEM64
ENSG00000282828	AC009971.1
ENSG00000236279	CLEC2L
ENSG00000161013	MGAT4B
ENSG00000246705	H2AJ
ENSG00000136231	IGF2BP3
ENSG00000196611	MMP1
ENSG00000283632	EXOC3L2
ENSG00000095739	BAMBI
ENSG00000163590	PPM1L
ENSG00000113083	LOX
ENSG00000198752	CDC42BPB
ENSG00000140830	TXNL4B
ENSG00000183454	GRIN2A
ENSG00000224116	INHBA-AS1
ENSG00000142949	PTPRF
ENSG00000245008	AP001122.1
ENSG00000172992	DCAKD
ENSG00000182747	SLC35D3
ENSG00000124302	CHST8
ENSG00000164171	ITGA2
ENSG00000285043	AC093512.2
ENSG00000287860	AC079804.3
ENSG00000254933	AP000785.1
ENSG00000010278	CD9
ENSG00000110799	VWF
ENSG00000169129	AFAP1L2
ENSG00000129354	AP1M2
ENSG00000159461	AMFR
ENSG00000129473	BCL2L2
ENSG00000064601	CTSA
ENSG00000166333	ILK
ENSG00000204872	NAT8B

ENSG00000255200	PGAM1P8
ENSG00000151748	SAV1
ENSG00000123191	ATP7B
ENSG00000134297	PLEKHA8P1
ENSG00000137225	CAPN11
ENSG00000168135	KCNJ4
ENSG00000184113	CLDN5
ENSG00000247735	AC120114.1
ENSG00000132613	MTSS2
ENSG00000099785	MARCHF2
ENSG00000125952	MAX
ENSG00000104067	TJP1
ENSG00000197147	LRRC8B
ENSG00000163833	FBXO40
ENSG00000100592	DAAM1
ENSG00000177697	CD151
ENSG00000176170	SPHK1
ENSG00000075945	KIFAP3
ENSG00000126856	PRDM7
ENSG00000243709	LEFTY1
ENSG00000117640	MTFR1L
ENSG00000162430	SELENON
ENSG00000255364	SMILR
ENSG00000185305	ARL15
ENSG00000128272	ATF4
ENSG00000165757	JCAD
ENSG00000023697	DERA
ENSG00000103769	RAB11A
ENSG00000137801	THBS1
ENSG00000162614	NEXN
ENSG00000078018	MAP2
ENSG00000279138	AP002847.1
ENSG00000184678	H2BC21
ENSG00000120903	CHRNA2
ENSG00000125388	GRK4
ENSG00000287514	AL450267.2
ENSG00000149932	TMEM219
ENSG00000229754	CXCR2P1
ENSG00000227550	TRBV7-5
ENSG00000170425	ADORA2B

ENSG00000260032	NORAD
ENSG00000255325	AC108136.1
ENSG00000237529	AL137847.2
ENSG00000106484	MEST
ENSG00000168734	PKIG
ENSG00000165091	TMC1
ENSG00000230749	MEIS1-AS2
ENSG00000108187	PBLD
ENSG00000270689	BUD13P1
ENSG00000253355	AP003469.1
ENSG00000166035	LIPC
ENSG00000100003	SEC14L2
ENSG00000204323	SMIM5
ENSG00000064961	HMG20B
ENSG00000229619	MBNL1-AS1
ENSG00000169224	GCSAML
ENSG00000272701	MESTIT1
ENSG00000279673	AC092919.3
ENSG00000170421	KRT8
ENSG00000116962	NID1
ENSG00000088280	ASAP3
ENSG00000050628	PTGER3
ENSG00000169504	CLIC4
ENSG00000066185	ZMYND12
ENSG00000166311	SMPD1
ENSG00000232725	U52111.1
ENSG00000204614	TRIM40
ENSG00000079156	OSBPL6
ENSG00000244300	GATA2-AS1
ENSG00000245281	AC124242.1
ENSG00000160188	RSPH1
ENSG00000042062	RIPOR3
ENSG00000185909	KLHDC8B
ENSG00000213625	LEPROT
ENSG00000177324	BEND2
ENSG00000250348	AC113404.1
ENSG00000130176	CNN1
ENSG00000186792	HYAL3
ENSG00000102572	STK24
ENSG00000254506	AP003080.1

ENSG00000255002	LINC02324
ENSG00000106868	SUSD1
ENSG00000127314	RAP1B
ENSG00000128311	TST
ENSG00000073464	CLCN4
ENSG00000239405	TMED10P2
ENSG00000125967	NECAB3
ENSG00000163734	CXCL3
ENSG00000253737	AP003469.3
ENSG00000168067	MAP4K2
ENSG00000113638	TTC33
ENSG00000236411	NDUFAF4P3
ENSG00000184838	PRR16
ENSG00000229666	MAST4-AS1
ENSG00000123500	COL10A1
ENSG00000203395	AC015969.1
ENSG00000273355	AP000894.4
ENSG00000151789	ZNF385D
ENSG00000105507	CABP5
ENSG00000117586	TNFSF4
ENSG00000040531	CTNS
ENSG00000257267	ZNF271P
ENSG00000274139	AC090164.2
ENSG00000122778	KIAA1549
ENSG00000254718	AL157756.1
ENSG00000011258	MBTD1
ENSG00000285909	AP002762.2
ENSG00000088836	SLC4A11
ENSG00000134030	CTIF
ENSG00000210151	MT-TS1
ENSG00000152128	TMEM163
ENSG00000248334	WHAMMP2
ENSG00000254810	AP001189.3
ENSG00000147862	NFIB
ENSG00000129636	ITFG1
ENSG00000284391	AL139398.1
ENSG00000222032	AC112721.2
ENSG00000174233	ADCY6
ENSG00000004776	HSPB6
ENSG00000163738	MTHFD2L

ENSG00000105499	PLA2G4C
ENSG00000039987	BEST2
ENSG00000143811	PYCR2
ENSG00000174640	SLCO2A1
ENSG00000233427	AL009181.2
ENSG00000259120	SMIM6
ENSG00000122970	IFT81
ENSG00000144868	TMEM108
ENSG00000169925	BRD3
ENSG00000271743	AF287957.1
ENSG00000124145	SDC4
ENSG00000163359	COL6A3
ENSG00000230385	AC012507.1
ENSG00000167414	GNG8
ENSG00000058866	DGKG
ENSG00000263155	MYZAP
ENSG00000132535	DLG4
ENSG00000198805	PNP
ENSG00000186480	INSIG1
ENSG00000092096	SLC22A17
ENSG00000142694	EVA1B
ENSG00000158164	TMSB15A
ENSG00000102893	PHKB
ENSG00000055813	CCDC85A
ENSG00000134853	PDGFRA
ENSG00000226191	CLK3P2
ENSG00000270055	AC127502.2
ENSG00000224846	NQO2-AS1
ENSG00000260249	AC007608.3
ENSG00000277639	AC007906.2
ENSG00000198695	MT-ND6
ENSG00000171385	KCND3
ENSG00000102755	FLT1
ENSG00000166340	TPP1
ENSG00000239445	ST3GAL6-AS1
ENSG00000270578	AP000787.2
ENSG00000175093	SPSB4
ENSG00000106211	HSPB1
ENSG00000184785	SMIM10
ENSG00000205795	CYS1

ENSG00000236671	PRKG1-AS1
ENSG00000236304	AP001189.1
ENSG00000137878	GCOM1
ENSG00000237419	AL954642.1
ENSG00000227199	ST7-AS1
ENSG00000166831	RBPMS2
ENSG00000256050	AL583722.1
ENSG00000184489	PTP4A3
ENSG00000255363	LINC02757
ENSG00000102468	HTR2A
ENSG00000183690	EFHC2
ENSG00000079257	LXN
ENSG00000187391	MAGI2
ENSG00000205456	TP53TG3D
ENSG00000171502	COL24A1
ENSG00000153904	DDAH1
ENSG00000101470	TNNC2
ENSG00000254717	GLYATL1P2
ENSG00000107186	MPDZ
ENSG00000143324	XPR1
ENSG00000154188	ANGPT1
ENSG00000272523	LINC01023
ENSG00000068615	REEP1
ENSG00000231437	LINC01750
ENSG00000213949	ITGA1
ENSG00000134201	GSTM5
ENSG00000233968	AL157895.2
ENSG00000242516	LINC00960
ENSG00000114805	PLCH1
ENSG00000133454	MYO18B
ENSG00000169908	TM4SF1
ENSG00000254332	AF201337.1
ENSG00000176732	PFN4
ENSG00000131095	GFAP
ENSG00000268320	SCGB1C2
ENSG00000210107	MT-TQ
ENSG00000177947	ODF3
ENSG00000181458	TMEM45A
ENSG00000078804	TP53INP2
ENSG00000187010	RHD

ENSG00000136295	TTYH3
ENSG00000188076	SCGB1C1
ENSG00000113924	HGD
ENSG00000223486	AC092198.1
ENSG00000103257	SLC7A5
ENSG00000134278	SPIRE1
ENSG00000259278	AC087878.1
ENSG00000284959	AC007262.2
ENSG00000279220	GPR1-AS
ENSG00000198075	SULT1C4
ENSG00000114654	EFCC1
ENSG00000207359	RNU6-925P
ENSG00000174951	FUT1
ENSG00000072135	PTPN18
ENSG00000235488	JARID2-AS1
ENSG00000228742	LINC02577
ENSG00000147400	CETN2
ENSG00000242611	AC093627.6
ENSG00000181322	NME9
ENSG00000100399	CHADL
ENSG00000134824	FADS2
ENSG00000177340	AC024940.1
ENSG00000205436	EXOC3L4
ENSG00000137507	LRRC32
ENSG00000175161	CADM2
ENSG00000212747	RTL8B
ENSG00000126458	RRAS
ENSG00000164066	INTU
ENSG00000102780	DGKH
ENSG00000265844	AC100863.1
ENSG00000153721	CNKSR3
ENSG00000261455	LINC01003
ENSG00000124006	OBSL1
ENSG00000270240	AC015849.1
ENSG00000144283	PKP4
ENSG00000151838	CCDC175
ENSG00000227640	SOX21-AS1
ENSG00000249242	TMEM150C
ENSG00000255028	AP003049.2
ENSG00000177096	PHETA2

ENSG00000206560	ANKRD28
ENSG00000284633	AL031590.1
ENSG00000188039	NWD1
ENSG00000226824	AC006001.2
ENSG00000117245	KIF17
ENSG00000176697	BDNF
ENSG00000250722	SELENOP
ENSG00000088367	EPB41L1
ENSG00000220201	ZGLP1
ENSG00000143194	MAEL
ENSG00000223935	LGALSL-DT
ENSG00000149243	KLHL35
ENSG00000268433	MTDHP3
ENSG00000198964	SGMS1
ENSG00000176014	TUBB6
ENSG00000196549	MME
ENSG00000236936	AL031005.1
ENSG00000167693	NXN
ENSG00000169291	SHE
ENSG00000277229	AC084781.1
ENSG00000107669	ATE1
ENSG00000210127	MT-TA
ENSG00000166848	TERF2IP
ENSG00000100280	AP1B1
ENSG00000210194	MT-TE
ENSG00000089169	RPH3A
ENSG00000258388	PPT2-EGFL8
ENSG00000087245	MMP2
ENSG00000220517	ASS1P1
ENSG00000237473	AC068759.1
ENSG00000165046	LETM2
ENSG00000256092	SBNO1-AS1
ENSG00000144331	ZNF385B
ENSG00000134352	IL6ST
ENSG00000254186	AC113414.1
ENSG00000255126	AP003064.1
ENSG00000268278	AC123912.3
ENSG00000113761	ZNF346
ENSG00000226800	CACTIN-AS1
ENSG00000272168	CASC15

ENSG00000137857	DUOX1
ENSG00000261033	AC005730.2
ENSG00000210144	MT-TY
ENSG00000210100	MT-TI
ENSG00000168306	ACOX2
ENSG00000287037	AC097709.1
ENSG00000186716	BCR
ENSG00000262873	AC127496.5
ENSG00000118946	PCDH17
ENSG00000125898	FAM110A
ENSG00000166153	DEPDC4
ENSG00000136315	AL355922.1
ENSG00000169760	NLGN1
ENSG00000100842	EFS
ENSG00000109272	PF4V1
ENSG00000167646	DNAAF3
ENSG00000150510	FAM124A
ENSG00000236242	MYO16-AS1
ENSG00000241318	WDR82P2
ENSG00000248275	TRIM52-AS1
ENSG00000167644	C19orf33
ENSG00000237854	LINC00674
ENSG00000204710	SPDYC
ENSG00000119042	SATB2
ENSG00000197632	SERPINB2
ENSG00000174697	LEP
ENSG00000259781	HMGB1P6
ENSG00000115507	OTX1
ENSG00000099337	KCNK6
ENSG00000224914	LINC00863
ENSG00000272434	AC137630.3
ENSG00000166165	CKB
ENSG00000103160	HSDL1
ENSG00000204740	MALRD1
ENSG00000153707	PTPRD
ENSG00000254659	LINC02715
ENSG00000244459	AC147067.1
ENSG00000196872	KIAA1211L
ENSG00000175175	PPM1E
ENSG00000235522	AC010978.1

ENSG00000154914	USP43
ENSG00000224786	CETN4P
ENSG00000267064	UXT-AS1
ENSG00000159307	SCUBE1
ENSG00000243969	AC073359.2
ENSG00000235531	MSC-AS1
ENSG00000132329	RAMP1
ENSG00000263823	AC009831.1
ENSG00000106603	COA1
ENSG00000235802	HCFC1-AS1
ENSG00000223553	SMPD4P1
ENSG00000259953	AL138756.1
ENSG00000122420	PTGFR
ENSG00000170955	CAVIN3
ENSG00000278611	ZNF426-DT
ENSG00000225778	PROSER2-AS1
ENSG00000272555	AC009974.1
ENSG00000130052	STARD8
ENSG00000107984	DKK1
ENSG00000152767	FARP1
ENSG00000184304	PRKD1
ENSG00000276566	IGKV1D-13
ENSG00000130529	TRPM4
ENSG00000198729	PPP1R14C
ENSG00000138413	IDH1
ENSG00000135929	CYP27A1
ENSG00000245112	SMARCA5-AS1
ENSG00000121005	CRISPLD1
ENSG00000278763	FAM27B
ENSG00000124406	ATP8A1
ENSG00000261342	AC006538.1
ENSG00000254400	AC091564.2
ENSG00000065833	ME1
ENSG00000049860	HEXB
ENSG00000277112	ANKRD20A21P
ENSG00000158055	GRHL3
ENSG00000246792	AC106038.1
ENSG00000198553	KCNRG
ENSG00000236528	AL033528.3
ENSG00000232316	LINC02518

ENSG00000173175	ADCY5
ENSG00000147647	DPYS
ENSG00000250535	STK19B
ENSG00000214142	RPL7P60
ENSG00000283639	MIR9500
ENSG00000143768	LEFTY2
ENSG00000276509	AC239799.1
ENSG00000144485	HES6
ENSG00000170891	CYTL1
ENSG00000272360	AC116036.2
ENSG00000227258	SMIM2-AS1
ENSG00000228570	NUTM2E
ENSG00000153237	CCDC148
ENSG00000100097	LGALS1
ENSG00000231636	AGBL5-AS1
ENSG00000152315	KCNK13
ENSG00000272153	AL365330.1
ENSG00000233270	SNRPEP4
ENSG00000164879	CA3
ENSG00000280064	AC130304.1
ENSG00000259353	AC090515.5
ENSG00000167767	KRT80
ENSG00000154133	ROBO4
ENSG00000008394	MGST1
ENSG00000128602	SMO
ENSG00000047662	FAM184B
ENSG00000139044	B4GALNT3
ENSG00000275597	ANAPC1P5
ENSG00000171962	DRC3
ENSG00000210176	MT-TH
ENSG00000272696	AL359091.3
ENSG00000186818	LILRB4
ENSG00000166073	GPR176
ENSG00000226251	LINC02608
ENSG00000174950	CD164L2
ENSG00000092850	TEKT2
ENSG00000175509	AL078621.2
ENSG00000211974	IGHV2-70D
ENSG00000204001	LCN8
ENSG00000100036	SLC35E4

ENSG00000115616 | SLC9A2

Online Table 13. GOBP pathway enrichment analysis for MEgreen module member genes.

module	geneset_id	description	p.adj	p.val	odds.ratio	n_fg	n_bg	fg_freq	bg_freq	n_set
green	GO:0007596	blood coagulation	7.74E-19	8.06E-22	6.13	53	272	0.0891	0.0187	431
green	GO:0050817	coagulation	1.18E-18	1.37E-21	6.04	53	275	0.0891	0.0189	439
green	GO:0007599	hemostasis	1.34E-18	1.63E-21	6.01	53	276	0.0891	0.019	438
green	GO:0030168	platelet activation	6.62E-17	8.76E-20	8.91	36	136	0.0605	0.00936	183
green	GO:0002576	platelet degranulation	1.11E-13	2.31E-16	8.99	29	108	0.0487	0.00743	131
green	GO:1903034	regulation of response to wounding	7.40E-10	3.14E-12	6.26	27	132	0.0454	0.00908	210
green	GO:0061041	regulation of wound healing	1.13E-09	5.15E-12	7.1	24	106	0.0403	0.00729	177
green	GO:0070527	platelet aggregation	1.73E-09	8.18E-12	12	17	51	0.0286	0.00351	68
green	GO:0003018	vascular process in circulatory system	4.09E-08	2.50E-10	5.99	23	116	0.0387	0.00798	201
green	GO:0034109	homotypic cell-cell adhesion	4.15E-08	2.56E-10	8.33	18	70	0.0303	0.00482	90
green	GO:0006937	regulation of muscle contraction	1.09E-07	7.42E-10	5.91	22	112	0.037	0.0077	209
green	GO:0030193	regulation of blood coagulation	2.52E-07	1.86E-09	8.53	16	61	0.0269	0.0042	103
green	GO:1900046	regulation of hemostasis	2.52E-07	1.86E-09	8.53	16	61	0.0269	0.0042	104
green	GO:0061045	negative regulation of wound healing	4.28E-07	3.31E-09	8.99	15	55	0.0252	0.00378	93
green	GO:0050818	regulation of coagulation	5.10E-07	3.99E-09	8	16	64	0.0269	0.0044	109
green	GO:0006939	smooth muscle contraction	6.48E-07	5.09E-09	7.83	16	65	0.0269	0.00447	120
green	GO:1903035	negative regulation of response to wounding	1.27E-06	1.03E-08	7.38	16	68	0.0269	0.00468	108
green	GO:0006940	regulation of smooth muscle contraction	2.32E-06	2.02E-08	11	12	38	0.0202	0.00261	68
green	GO:0030195	negative regulation of blood coagulation	2.32E-06	2.02E-08	11	12	38	0.0202	0.00261	68
green	GO:1900047	negative regulation of hemostasis	2.32E-06	2.02E-08	11	12	38	0.0202	0.00261	69

green	GO:0090257	regulation of muscle system process	2.65E-06	2.38E-08	4.2	25	169	0.042	0.0116	324
green	GO:0050819	negative regulation of coagulation	4.19E-06	3.87E-08	10.2	12	40	0.0202	0.00275	73
green	GO:0090287	regulation of cellular response to growth factor stimulus	6.69E-06	6.36E-08	3.84	26	190	0.0437	0.0131	356
green	GO:0035150	regulation of tube size	5.18E-05	5.57E-07	5.25	16	89	0.0269	0.00612	166
green	GO:0035296	regulation of tube diameter	5.18E-05	5.57E-07	5.25	16	89	0.0269	0.00612	165
green	GO:0097746	regulation of blood vessel diameter	5.18E-05	5.57E-07	5.25	16	89	0.0269	0.00612	165
green	GO:0097581	lamellipodium organization	7.86E-05	8.86E-07	5.87	14	71	0.0235	0.00488	97
green	GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	9.00E-05	1.03E-06	3.7	22	165	0.037	0.0114	330
green	GO:0007179	transforming growth factor beta receptor signaling pathway	0.000132	1.54E-06	3.74	21	156	0.0353	0.0107	270
green	GO:0030100	regulation of endocytosis	0.000272	3.49E-06	3.53	21	164	0.0353	0.0113	233
green	GO:0071560	cellular response to transforming growth factor beta stimulus	0.000321	4.18E-06	3.26	23	193	0.0387	0.0133	327
green	GO:0071559	response to transforming growth factor beta	0.000435	5.92E-06	3.18	23	197	0.0387	0.0136	336
green	GO:0019935	cyclic-nucleotide-mediated signaling	0.00044	6.02E-06	4.03	17	118	0.0286	0.00812	283
green	GO:0030198	extracellular matrix organization	0.000443	6.08E-06	2.86	27	255	0.0454	0.0175	473
green	GO:0045987	positive regulation of smooth muscle contraction	0.000443	6.09E-06	13.8	7	19	0.0118	0.00131	34
green	GO:0043062	extracellular structure organization	0.000472	6.54E-06	2.85	27	256	0.0454	0.0176	476
green	GO:0031589	cell-substrate adhesion	0.000598	8.67E-06	2.74	28	275	0.0471	0.0189	430
green	GO:0007160	cell-matrix adhesion	0.00066	9.77E-06	3.28	21	175	0.0353	0.012	267
green	GO:0007188	adenylate cyclase-modulating G protein-coupled receptor signaling pathway	0.000706	1.06E-05	3.84	17	123	0.0286	0.00846	304

green	GO:0007187	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	0.00072	1.09E-05	3.65	18	136	0.0303	0.00936	361
green	GO:0043552	positive regulation of phosphatidylinositol 3-kinase activity	0.000743	1.13E-05	9.48	8	28	0.0134	0.00193	34
green	GO:0038084	vascular endothelial growth factor signaling pathway	0.000946	1.50E-05	9.03	8	29	0.0134	0.00199	52

COMBAT Consortium members and affiliations

David J Ahern (2), Zhichao Ai (2), Mark Ainsworth (19), Chris Allan (1), Alice Allcock (1), Brian Angus (19), M Azim Ansari (1,6), Carolina V Arancibia-Cárcamo (16, 18), Dominik Aschenbrenner (16), Moustafa Attar (1,2), J Kenneth Baillie (28), Eleanor Barnes (6,16,18,19), Rachael Bashford-Rogers (1), Archana Bashyal (19), Sally Beer (39), Georgina Berridge (5), Amy Beveridge (8), Sagida Bibi (8), Tihana Bicanic (33), Luke Blackwell (8), Paul Bowness (13,18), Andrew Brent (19,38), Andrew Brown (1), John Broxholme (1), David Buck (1), Katie L Burnham (27), Helen Byrne (7), Susana Camara (8), Ivan Candido Ferreira (3), Philip Charles (4,5), Wentao Chen (20), Yi-Ling Chen (3), Amanda Chong (1), Elizabeth A Clutterbuck (8), Mark Coles (2), Christopher P Conlon (17,19,31), Richard Cornall (25), Adam P Cribbs (13), Fabiola Curion (1), Emma E Davenport (27), Neil Davidson (19), Simon Davis (5), Calliope A Dendrou (1), Julie Dequaire (19), Lea Dib (2), James Docker (1), Christina Dold (8), Tao Dong (17,25), Damien Downes (3), Hal Drakesmith (3), Susanna J Dunachie (6,19,25,31), David A Duncan (3,21), Chris Eijsbouts (1,4), Robert Esnouf (1,4), Alexis Espinosa (39), Rachel Etherington (3,25), Benjamin Fairfax (3,10,18,19), Rory Fairhead (19), Hai Fang (1), Shayan Fassih (19), Sally Felle (8), Maria Fernandez Mendoza (39), Ricardo Ferreira (1), Roman Fischer (5,17), Thomas Foord (19), Aden Forrow (7), John Frater (6,19), Anastasia Fries (19,20), Veronica Gallardo Sanchez (19), Lucy C Garner (3,16), Clementine Geeves (1), Dominique Georgiou (39), Leila Godfrey (1), Tanya Golubchik (1,4), Maria Gomez Vazquez (2), Angie Green (1), Hong Harper (1), Heather A Harrington (1,7), Raphael Heilig (5), Svenja Hester (5), Jennifer Hill (8), Charles Hinds (34), Clare Hird (19), Ling-Pei Ho (20,25), Renee Hoekzema (7), Benjamin Hollis (12), Jim Hughes (3,26), Paula Hutton (19), Matthew A Jackson-Wood (2), Ashwin Jainarayanan (2), Anna James-Bott (13), Kathrin Jansen (2), Katie Jeffery (19), Elizabeth Jones (8), Luke Jostins (2,4), Georgina Kerr

(36), David Kim (19), Paul Klenerman (6,16), Julian C Knight (1,17,18,19), Vinod Kumar (2), Piyush Kumar Sharma (3,10), Prathiba Kurupati (3,25), Andrew Kwok (1), Angela Lee (1), Aline Linder (8), Teresa Lockett (19), Lorne Lonie (1), Maria Lopopolo (1), Martyna Lukoseviciute (3), Jian Luo (20), Spyridoula Marinou (8), Brian Marsden (2,22), Jose Martinez (39), Philippa C Matthews (6), Michalina Mazurczyk (3,25), Simon McGowan (3), Stuart McKechnie (19), Adam Mead (3,18,19), Alexander J Mentzer (1,19), Yuxin Mi (1), Claudia Monaco (2), Ruddy Montadon (1), Giorgio Napolitani (3,25), Isar Nassiri (3,10), Alex Novak (39), Darragh P O'Brien (5), Daniel O'Connor (8,18), Denise O'Donnell (19), Graham Ogg (3,18,25), Lauren Overend (1), Inhye Park (2), Ian Pavord (20), Yanchun Peng (3,17,25), Frank Penkava (13), Mariana Pereira Pinho (3,25), Elena Perez (39), Andrew J Pollard (8,18), Fiona Powrie (2), Bethan Psaila (3,18), T Phuong Quan (32), Emmanouela Repapi (3,26), Santiago Revale (1), Laura Silva-Reyes (8), Jean-Baptiste Richard (2,3), Charlotte Rich-Griffin (1), Thomas Ritter (19), Christine S Rollier (8), Matthew Rowland (19), Fabian Ruehle (11), Mariolina Salio (3,25), Stephen Nicholas Sansom (2), Raphael Sanches Peres (2), Alberto Santos Delgado (4,23,24), Tatjana Sauka-Spengler (3), Ron Schwessinger (3), Giuseppe Scozzafava (1), Gavin Screaton (1,36), Anna Seigal (7), Malcolm G Semple (29), Martin Sergeant (3), Christina Simoglou Karali (3), David Sims (3), Donal Skelly (6,14,19), Hubert Slawinski (1), Alberto Sobrinodiaz (39), Nikolaos Sousos (3,18,19), Lizzie Stafford (19), Lisa Stockdale (8), Marie Strickland (3), Otto Sumray (7,37), Bo Sun (1), Chelsea Taylor (3,10), Stephen Taylor (3), Adan Taylor (19), Supat Thongjuea (3), Hannah Thraves (39), John A Todd (1), Adriana Tomic (8), Orion Tong (3,10), Amy Trebes (1), Dominik Trzupek (1), Felicia Anna Tucci (1), Lance Turtle (29,30), Irina Udalova (2), Holm Uhlig (16,19), Erinke van Grinsven (2), Iolanda Vendrell (5), Marije Verheul (8), Alexandru Voda (2), Guanlin Wang (3), Lihui Wang (2), Dapeng Wang (1), Peter Watkinson (14,19), Robert Watson (3,10,19), Michael

Weinberger (3,35), Justin Whalley (1), Lorna Witty (1), Katherine Wray (3,19), Luzheng Xue (20), Hing Yuen Yeung (1), Zixi Yin (17,25), Rebecca K Young (19), Jonathan Youngs (33), Ping Zhang (1), Yasemin-Xiomara Zurke (2).

1: Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, UK

2: Kennedy Institute for Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Diseases, University of Oxford, UK

3: MRC Weatherall Institute of Molecular Medicine (WIMM), University of Oxford, UK

4: Big Data Institute, Nuffield Department of Medicine, University of Oxford, UK

5: Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, UK

6: Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, UK

7: Mathematical Institute, University of Oxford, UK

8: Oxford Vaccine Group, Department of Paediatrics, University of Oxford, UK

9: Department of Paediatrics, University of Oxford, UK

10: Department of Oncology, University of Oxford, UK

11: Department of Physics, University of Oxford, UK

12: Jenner Institute, Nuffield Department of Medicine, University of Oxford, UK

13: Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Diseases, University of Oxford, UK

14: Nuffield Department of Clinical Neuroscience, University of Oxford, UK

15: Department of Statistics, University of Oxford, UK

16: Translational Gastroenterology Unit, University of Oxford, UK

17: Chinese Academy of Medical Science Oxford Institute (COI), University of Oxford, UK

18: NIHR Oxford Biomedical Research Centre, Oxford, UK

19: Oxford University Hospitals NHS Foundation Trust, Oxford, UK

20: Respiratory Medicine Unit, Nuffield Department of Medicine, University of Oxford, UK

21: Diamond Light Source, Didcot, UK

22: Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, UK

23: Center for Health Data Science, University of Copenhagen, DK

24: Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, DK

25: MRC Human Immunology Unit, Radcliffe Department of Medicine, University of Oxford, UK

26: MRC WIMM Centre for Computational Biology, Radcliffe Department of Medicine, University of Oxford, UK

27: Wellcome Sanger Institute, Cambridge, UK

28: Genetics and Genomics, Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, UK

29: NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool

30: Tropical and Infectious Diseases Unit, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK

31: Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, UK

32: Experimental Medicine Division, Nuffield Department of Medicine, University of Oxford, UK

33: Institute for Infection and Immunity, St George's University of London, UK

34: The William Harvey Research Institute, Bart’s and The London, Queen Mary University of London, UK

35: Department of Physiology, Anatomy and Genetics, University of Oxford, UK

36: Medical Sciences Division, University of Oxford, UK

37: Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, UK

38: Nuffield Department of Medicine, University of Oxford, UK

39: Emergency Medicine Research Oxford, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

ORFAN Investigators and Affiliations

Adrian P Banning (1), Alexios S Antonopoulos (1,2), Amrita Bajaj (3), Andrew David Kelion (1), Aparna Deshpande (3), Attila Kardos (4), Benjamin Hudson (5), Bon-Kwon Koo (6), Cheerag Shirodaria (7), Cheng Xie (1), Christos Kotanidis (1), Ciara Mahon (8), Colin Berry (9), David Adlam (3), David Ernest Newby (10), Derek Leslie Connolly (11), Diane Elizabeth Scaletta (4), Donna Alexander (3), Ed Nicol (8), Elisa McAlindon (12), Evangelos Oikonomou (13), Francesca Pugliese (14), Gianluca Pontone (15), Giulia Benedetti (16), Guo-Wei He (17), Henry West (1), Hidekazu Kondo (1), Imre S Benedek (18), Intrajeet Das (3), John Deanfield (19), John Graby (5), John Greenwood (20), Jonathan Carl Luis Rodrigues (5), Junbo Ge (21), Keith M Channon (1), Larissa Fabritz (22), Li-Juan Fan (17), Lucy Kingham (1), Marco Guglielmo (15), Maria Lyasheva (1), Matthias Schmitt (23), Meinrad Beer (24), Michelle Louise Anderson (20), Milind Y Desai (25), Mohamed Marwan (26), Naohiko Takahashi (27), Nehal N Mehta (28), Neng Dai (21), Nicholas Screaton (29), Nikant Kumar Sabharwal (7), Pál Maurovich-Horvat (30), Praveen PG Rao (3), Rafail Angelos Kotronias (1), Rajesh Kumar Kharbanda (7), Rebecca Louise Preston (16), Richard James Wood (5), Ron Blankstein (31), Ronak Rajani (16), Saeed Mirsadraee (8), Shahzad Mahmood Munir (12), Sheena Thomas (1), Stefan Neubauer (1), Steffen Christoph Klömpken (24), Steffen E Petersen (14), Stephan Achenbach (26), Susan Anthony (7), Sze Mun Mak (16), Tarun Mittal (8), Theodora M Benedek (18), Vinoda Sharma (11), Wen-Hua Lin (17).

1: Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford

2: Hippokration General Hospital, Athens, Greece

3: University Hospitals of Leicester NHS Trust

4: Milton Keynes University Hospital NHS Foundation Trust

5: Royal United Hospitals Bath NHS Foundation Trust

6: Seoul National University, Seoul, South Korea

7: Oxford University Hospitals NHS Foundation Trust

8: Royal Brompton and Harefield Hospitals

9: BHF Glasgow Cardiovascular Research Centre, University of Glasgow

10: BHF Edinburgh Clinical Research Facility

11: Sandwell & West Birmingham Hospitals NHS Trust

12: Heart and Lung Centre, New Cross Hospital, Wolverhampton

13: Section of Cardiovascular Medicine. Yale-New Haven Hospital, Yale School of Medicine

14: NIHR Barts Cardiovascular Biomedical Research Centre

15: Centro Cardiologico Monzino IRCCS, University of Milan, Italy

16: Guy's and St Thomas' NHS Foundation Trust

17: TEDA International Cardiovascular Hospital, Tianjin, China

18: University of Medicine and Pharmacy of Tirgu Mures, Romania

19: University College London

20: Leeds Teaching Hospitals NHS Foundation Trust

21: Fudan University, China

22: University Hospitals Birmingham NHS Trust

23: University Hospital of Manchester Foundation Trust

24: University Hospital Ulm, Germany

25: Heart and Vascular Institute, Cleveland Clinic

26: Department of Cardiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

27: Oita University, Japan

28: National Institutes of Health, National Heart, Lung, and Blood Institute

29: Royal Papworth Hospital NHS Trust, Cambridge

30: Department of Radiology, MTA-SE Cardiovascular Imaging Research Group, Budapest, Hungary

31: Harvard Medical School, Boston, USA

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2-3
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	6-7
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7-9
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7-9
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7-9
	5b	Describe eligibility criteria for participants.	App 1
	5c	Give details of treatments received, if relevant.	N/A
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7
	6b	Report any actions to blind assessment of the outcome to be predicted.	App 6
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	App 13
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	App 6
Sample size	8	Explain how the study size was arrived at.	App 12
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	App 12
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	App 13
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	App 10-11
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	App 10-11
Risk groups	11	Provide details on how risk groups were created, if done.	App 13
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	App 17
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	App 31, 34-36
Model development	14a	Specify the number of participants and outcome events in each analysis.	App 18
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	App 39
	15b	Explain how to the use the prediction model.	12-14
Model performance	16	Report performance measures (with CIs) for the prediction model.	Figure 4
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	20-21
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	18
Implications	20	Discuss the potential clinical use of the model and implications for future research.	21
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	22
Funding	22	Give the source of funding and the role of the funders for the present study.	11

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.