Precision medicine for pandemics: stratification of COVID-19 molecular phenotypes defined by topological analysis of global blood gene expression.

Rebekah Penrice-Randal [1,2], Fabio Strazzeri [1], Benoit Ernst [3], Brice Van Eeckhout [4], Julien Guiot [3], Anna Julie Peired [5], Cosimo Nardi [5], Erika Parkinson [1],Monique Henket [3], Alicia Staderoli [3], Elora Guglielmi [3], Anne-Françoise Dive [3], Laurie Giltay [3], Sara Tomassetti [6], Rebecca Baker [7], Kit Howard [7], Catherine Hartley [2], Tessa Prince [2], Thomas Kleyntssens [4], Tommaso Manciulli [8], Ratko Djukanovic [9,10], Tristan Clark [9,10], Diana Baralle [10,11], Scott S Wagers [12], Xiaodan Xing [13], Yang Nan [13], Shiyi Wang [13], Simon Walsh [14] , Guang Yang [13], Paul J Skipp [1,15], Julian A Hiscox [2,16,17] , James P R Schofield [1].

Affiliations:

[1]TopMD Precision Medicine Ltd, Southampton, United Kingdom

[2]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK.

[3]Department of Respiratory Medicine, University Hospital of Liège, Liège, Belgium.

[4]Comunicare Solutions, Liège, Belgium.

[5]Department of Experimental and Clinical Biomedical Sciences “Mario Serio”, University of Florence, Viale Morgagni 50, 50134 Florence, Italy.

[6]Department of Clinical and Experimental Medicine, Careggi University Hospital, Largo Brambilla 3, 50134 Florence, IT

[7]Clinical Data Interchange Standards Consortium (CDISC), Austin, TX, United States of America

[8]Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy.

[9]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[10]National Institute for Health Research Southampton Biomedical Research Centre, Southampton, United Kingdom

[11]School of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[12]BioSci Consulting, Maasmechelen, Belgium

[13]Department of Bioengineering and Imperial-X, Imperial College London

[14]National Heart and Lung Institute, Imperial College London

[15]School of Biological Sciences, University of Southampton, Southampton, UK

[16]NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool, UK.

[17]A*STAR Infectious Diseases Laboratories (A*STAR ID Labs), Agency for Science, Technology and Research (A*STAR), Singapore.

Abstract:

Precision medicine offers a promising avenue for better therapeutic responses to pandemics such as COVID-19. This study leverages independent patient cohorts in Florence and Liège gathered under the umbrella of the DRAGON consortium for the stratification of molecular phenotypes associated with COVID-19 using topological analysis of global blood gene expression. Whole blood from 173 patients was collected and RNA was sequenced on the Novaseq platform. Molecular phenotypes were defined through topological analysis of gene expression relative to the biological network using the TopMD algorithm. The two cohorts from Florence and Liège allowed for independent validation of the findings in this study. Clustering of the topological maps of differential pathway activation revealed three distinct molecular phenotypes of COVID-19 in the Florence patient cohort, which were also observed in the Liège cohort.

Cluster 1 was characterised by high activation of pathways associated with ESC pluripotency, NRF2, and TGF-β receptor signalling. Cluster 2 displayed high activation of pathways including focal adhesion-PI3K-Akt-mTOR signalling and type I interferon induction and signalling, while Cluster 3 exhibited low IRF7-related pathway activation. TopMD was also used with the Drug-Gene Interaction Database (DGIdb), revealing pharmaceutical interventions targeting mechanisms across multiple phenotypes and individuals.

The data illustrates the utility of molecular phenotyping from topological analysis of blood gene expression, and holds promise for informing personalised therapeutic strategies not only for COVID-19 but also for Disease X. Its potential transferability across multiple diseases highlights the value in pandemic response efforts, offering insights before large-scale clinical studies are initiated.

Introduction:

The ongoing challenges of COVID-19, triggered by the emergence of SARS-CoV-2, necessitate a detailed understanding of disease heterogeneity. Despite extensive research characterising the host response to SARS-CoV-2 through pre-clinical (1, 2), and clinical (3-6) functional genomic data, there have been limited approaches that have used data from and encompassed the range of symptom severity, disease heterogeneity and delivered personalised medicine.

Examination of gene expression patterns in blood has been used in previous studies to identify molecular phenotypes associated with different disease profiles in several emerging viral infections including Ebola virus (EBOV) (7) and SARS-CoV-2 (1, 2, 4, 5), as well as more endemic infections such as influenza virus (8). Medical countermeasures focus on either reducing viral load through anti-virals. These target viral biology or modulate the host response to infection to reduce sequalae such as inflammation. For many viruses there is a clear correlation between viral load, disease severity and outcome (survival/death). This is best typified by the Ebola virus where low viral loads correlate with survival and high viral loads correlate with death (9). For SARS-CoV-2 this correlation is less obvious. In animal models of disease, such as the ferret, viral load was correlated with symptomology (10); in humans, there is less data to support an association between viral load and disease. However, studies have shown that severe COVID-19 is associated with dysregulated immune pathology in organs such as the lungs and the respiratory tract (3, 11).

With any emerging viral pathogen, direct acting antivirals take time to develop and trial. Identifying therapeutics that can modulate the host response to reduce symptomology remain a priority. Being able to rapidly characterise aberrations in host pathways that lead to disease and marrying this with therapeutics on the FDA approved list will enhance pandemic preparedness and rapid response. Therefore, a deeper understanding of the host response can be used to guide the selection of host directed medication countermeasures.

The field of digital health and precision medicine is rapidly evolving, with emerging technologies and initiatives aimed at integrating diverse datasets to inform clinical decision-making. In this study we offer a novel way to analyse complex data collected by the DRAGON international consortium which enables rapid identification of targets for treatment by novel and/or re-purposed drugs. Within DRAGON, efforts have been made to harmonise data in digital healthcare, proposing guidelines for the integration of clinical data from various modalities. (12). Additionally, an online platform has been developed to host validated COVID-19 predictive models, facilitating their utilisation by clinicians in real-time decision-making (13). However, challenges persist, as evidenced by the limited success of outcome prediction models for COVID-19 patients based on demographic and comorbidity data, which highlights the need for more sophisticated approaches (14).

While omics data has been instrumental in advancing our understanding of SARS-CoV-2 and COVID-19, its integration into digital health platforms for clinical decision-making remains limited (15-17). Traditional molecular phenotyping approaches often provide only shallow insights. In previous work, using topological analysis, we demonstrated how gene expression data derived from whole blood at the time of admission could predict ICU admission (5). However, the current study analysed the blood transcriptomes of patients with COVID-19 as part of the DRAGON-EU consortium and used TopMD, an algorithm that considers all available data across a landscape of pathways, to characterize molecular phenotypes of COVID-19 patients admitted to hospital. Pathways were identified that correlated with clinical disease in the patient cohort. TopMD mapped pathways onto a database containing information on FDA approved drugs and their known gene and pathway interactions to generate a list of potential therapeutics for modulating severe COVID-19. The ability to rapidly identify and therapeutically modulate host pathways responsible for disease with pre-existing medical countermeasures will be important in the emergence of novel diseases and future pandemics.

This study describes an analysis of the blood transcriptomes of patients with COVID-19 admitted to hospital in Liège and Florence between February and July 2021, as part of the DRAGON-EU consortium. Alongside collecting blood samples, demographic and clinical observations were recorded; additionally, CT scan data were

obtained for a subset of these patients. We applied an unsupervised approach, in which we characterised the molecular phenotypes of patients within this cohort. We have previously reported the development of a gene signature in patients with COVID-19, predictive of admission to ICU (5). This predictive signature revealed the activation of pathways regulating epidermal growth factor receptor (EGFR) signalling, peroxisome proliferator-activated receptor alpha (PPAR-α) signalling and transforming growth factor beta (TGF-β) signalling. The observed molecular phenotype aligns with the mechanisms implicated in pulmonary fibrosis, which is also associated with increased severity of disease (18-20).

164 **Methods:**

165

166 **Study population and sample collection and ethics**

167 Blood samples were obtained from 132 patients severe enough to require admission
168 because of COVID-19 at Careggi University Hospital, Florence, Italy, and 41 from a
169 pre-defined, separate patient cohort in Liège, between February and July 2021. All
170 patients tested positive nasopharyngeal swab PCR for SARS-CoV-2 infection. Blood
171 samples were collected on Day 0 of hospital admission. The protocol was approved
172 by the ethics committee of the University Hospital of Liège (reference number 2021/89)
173 and the ethics committee of the UNIFI (#18085/OSS). Informed consent was obtained
174 for every participant.

175

176 **Ethical Approval statement**

177 The work described has been carried out in accordance with The Code of Ethics of
178 the World Medical Association (Declaration of Helsinki) for experiments involving
179 humans. All procedures were performed in compliance with relevant laws and
180 institutional guidelines and have been approved by the appropriate institutional
181 committees. Informed consent was obtained for every participant.

182 Clinical data were collected from the patients' electronic medical records by the
183 investigators, and included age, sex, BMI, comorbid conditions etc. The data were
184 then assembled using the Study Data Tabulation Model (SDTM) data format
185 developed by the Clinical Data Interchange Standards Consortium (CDISC).

186 **Chest CT analysis**

187 Out of the 173 patients with RNA sequencing data, chest CT data was obtained from
188 109 patients using a 128-detector multislice Spiral Computed Tomography (MSCT)
189 (Somatom Definition AS, Siemens Healthcare, Erlangen, Germany) applying the
190 following parameters: current × exposure time 150 mAs, tube voltage 100 kV, rotation
191 time 0.3 s, pitch 1.2 mm, pixel size 0.465 mm, beam collimation 128 × 0.6 mm, both
192 slice thickness and reconstruction 1 mm, and reconstruction kernel Bf70 very sharp.
193 Axial images were carried out from lung apexes to bases with patient at full inspiration
194 mand breath hold. Post-processing, 1-mm-thick sections were reconstructed on
195 coronal and sagittal planes oriented on the tracheal plane. Intravenous contrast

medium was not administered. Chest CT images were displayed on a 24-inch medical monitor with a 3-megapixel Barco display (Barco, Kortrijk, Belgium) and 2048 x 1536 resolution. The software programs originally implemented to MSCT were used for image assessment. Images of each patient were evaluated for scan quality considering inspiratory level and motion artifacts. Data pulled out from CT examinations included CO-RADS, chest CT score, dominant pattern, and typical/atypical findings. Specifically:

CO-RADS

CO-RADS score based on COVID-19 lung involvement and variable from 1 to 5, with higher values reflecting a greater level of suspicion of COVID-19 infection with lung involvement. CO-RADS is a score used to diagnose COVID-19 and does not inevitably reproduce the severity of lung alterations. Low scores corresponded to CT examinations with alterations less likely related to COVID-19 infection. The 5-score CO-RADS scale is as follows: 1: very low level of suspicion; 2: low level of suspicion; 3: equivocal findings; 4: high level of suspicion; 5: very high level of suspicion.

Chest CT score for lobe involvement

Ranging from 0 to 5, namely 0: 0%; 1: <5%; 2: 5-25%; 3: 26-50%; 4: 51-75%; 5: >75%.

Dominant chest CT pattern

Evaluated in relation to the prevalent alterations among ground-glass opacities, consolidations, ground-glass opacities together with consolidations, crazy-paving, and reverse halo, as defined by the Fleischner Society.

Dominant chest CT distribution

Lower lobes, upper lobes, peripheral, bronchocentric, dorsal, or diffuse.

Additional COVID-19 related findings

Represented by pleural thickening, vascular enlargement, subpleural sign, halo sign air, bubble sign, perilobular pattern, and subpleural sparing.

Additional findings not typical for COVID-19

Represented by pleural effusion, pericardial effusion, lymphadenopathy,

226 cavitation, tree-in-bud, discrete small nodules, isolated lobar/segmental
227 consolidation, atelectasis, and smooth interlobular septal thickening.

228 **RNA extraction**

229 Total RNA was extracted from PAXgene BRT using the PAXgene Blood RNA Kit
230 (PreAnalytix), according to the manufacturer's protocol. Extracted RNA was stored at
231 -80°C until further use. Following the manufacturer's protocols, total RNA was used
232 as input material into the QIAseq FastSelect–rRNA/Globin Kit (Qiagen) protocol to
233 remove cytoplasmic and mitochondrial rRNA and globin mRNA with a fragmentation
234 time of 7 or 15 minutes. Subsequently the NEBNext® Ultra™ II Directional RNA Library
235 Prep Kit for Illumina® (New England Biolabs) was used to generate the RNA libraries,
236 followed by 11 or 13 cycles of amplification and purification using AMPure XP beads.
237 Each library was quantified using Qubit and the size distribution assessed using the
238 Agilent 2100 Bioanalyser and the final libraries were pooled in equimolar ratios.
239 Libraries were sequenced using 150 bp paired-end reads on an Illumina® NovaSeq
240 6000 (Illumina®, San Diego, USA).
241

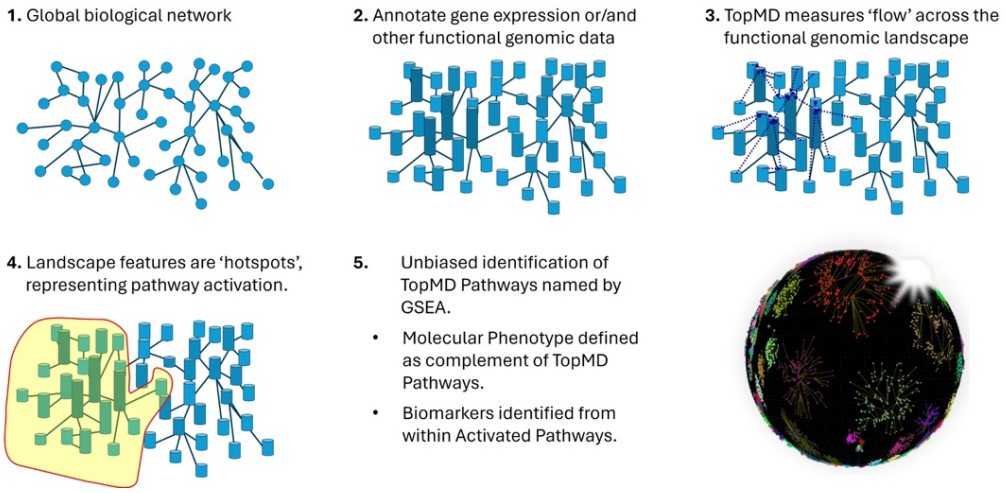
242 **Bioinformatics**

243 Raw fastq files were trimmed using fastp (21). Trimmed paired end sequencing reads
244 were inputted into salmon (v1.5.2) using the -l A –validateMappings –SeqBias –gcBias
245 parameters (22). Quant files generated with salmon were imported into RStudio (4.1.1)
246 using tximport to infer gene expression (23). The edgeR package (3.34.1) was used
247 to normalise and scale sequencing libraries (24). Sequencing reads are available
248 under BioProject ID: PRJNA1085259 on Short Read Archive (SRA).
249

250 **Molecular phenotypes mapped by topological analysis**

251 Molecular phenotypes were mapped by topological analysis, using TopMD to measure
252 the shape of global gene expression relative to the biological network (TopMD Patent
253 number GB202306368D0). TopMD works in the following way: The biological network
254 used was an interaction network retrieved from the STRING database (25).The gene
255 nodes of the biological network were assigned vertices according to the measured
256 gene expression. The topological shape, or landscape, of this network is then
257 measured by TopMD's algorithm, clustering differential gene expression hotspots,
258 corresponding to modulated gene pathways. These pathways have 'volume'

comprising the sum of squared differential gene expression of clustered genes, where the most differentially activated pathways have the highest pathway topological volumes. The molecular phenotype is defined as the global profile of volumes of differential pathway activation.



Drug interactions mapped by topological analysis

Due to the power of TopMD analysis we can group genes depending on their expression values, this means that for each average expression of any cluster of samples, and even on individual samples, we can extrapolate a tailored gene set of activated gene-groups for such expression. These gene-groups can be then compared to other gene sets, as in GSEA, as well against genes activated by specific drugs. To do so, we utilised the Drug-Gene Interaction Database (26) obtained using genes or gene products that are known or predicted to interact with drugs, and compared via a binomial distribution test, the probability that an overlap between such genes and a TopMD gene-group was random. This was measured using a p-value associated with binomial statistic, together with other measures, such as the (Bonferroni) adjusted p-value, a TopMD volume (combining volume of the shape with the statistical significance of the drug-group combination) as well as an activation value, sum of the Log2 fold-change of those genes belonging to both the drug associated gene set and the TopMD group.

Regression

Regression analysis was carried out using a Logistic regression model with the following optimisation problem:

$$\min_{w} C \sum_{i=1}^{n} s_i \left(-y_i \log(\hat{p}(X_i)) - (1 - y_i) \log(1 - \hat{p}(X_i))\right) + r(w)$$

Where X is the **pathway matrix** and y is the vector of the classification, 0 when the i-th sample is in the class considered and 1 otherwise. We considered a regularisation parameter C value of 1. For the penalisation term *r(w)* for the regression weights w, we considered an ElasticNet penalisation with the **l1 ratio** parameter value of 0.5

$$\frac{1-\rho}{2} w^T w + \rho \|w\|_1$$

The probability the i-th sample with **pathways values** equal to Xi is then:

$$\hat{p}(X_i) = \operatorname{expit}(X_i w + w_0) = \frac{1}{1 + \exp(-X_i w - w_0)}$$

With w0 the intercept. The python module used was **scikit-learn (version 1.4.1)** and the algorithm used LogisticRegression function in the linear_model submodule.

We performed a 70/30 balanced split in the data from both cohorts separately (?), with 10 different splits. For each class we performed the regression based on a different number of pathways, from 1 to 20, ranked in each split separately by their pathway volume. For each regression model so obtained an average score of both training and test splits was carried and the best model was selected using a max-min approach, that is the best model was the one with highest value min(AUC on Train, AUC on Test), to avoid selecting models which were ill-performing on train splits, but instead for random effects very well on test splits.

Patient Clustering

Pathway volumes were plotted on a PCA using PCAtools (v2.14.0), revealing 3 distinct clusters, confirmed by K-means clustering, based on pathway activation against healthy controls. The top ten (10%) of the PCA loadings were then extracted to identify

which pathways were driving cluster separation. To analyse differentially activated pathways between patient clusters, we calculated the average volume, across each cluster, of each pathway relative to the average of all the COVID-19 patients.

Logistic Regression Receiver Operating Characteristic (LRROC) analysis using patient clusters derived from the patient pathway volume matrix.

The area under the ROC curve (AUC) is a measure of the model's ability to distinguish between classes. A higher AUC indicates better discrimination and, consequently, stronger patient clusters. LRROC for Florence Patients: LRROC analysis was performed exclusively for Florence patients. The patient pathway volume matrix for Florence patients was utilized to train the LRROC model. The output consisted of a Receiver Operating Characteristic (ROC) curve, which depicted the classification performance of patient clusters based on pathway volume. To evaluate the model's generalization capability, the dataset was split into training and testing sets, and separate ROC curves were generated for each.

Validation of clusters for Liège Patients: The LRROC model trained on Florence patients was validated on Liège patients' data. Using the trained model, an additional ROC curve was generated solely for Liège patients to assess the model's performance in classifying Liège patient clusters based on pathway volume.

Integration into digital health platform

As a proof of concept, transcriptomics data and TopMD analysis were integrated with a healthcare platform ran by Comunicare (27). This was to highlight the possibilities of integrating omics data into healthcare and digital health platforms. Similar regression analysis of COVID-19 blood transcriptomes, predicting ICU admission, performed within the DRAGON scope (5) generated a linear model which is currently used to generate prediction scores between 0 and 1, using TopMD analysis of each sample submitted. In this way we can present TopMD analysis of individual samples compared to a healthy baseline, which includes pathway activation information, together with a similarity score to the ICU admitted average patient we extracted from previous data.

346 **Results:**

347

348 To investigate whether blood transcriptomic analysis coupled with a machine learning
349 approach underpinned by TopMD could be integrated with clinical data, RNA
350 sequencing was performed on peripheral blood obtained from 173 patients from Liège
351 (n=41) and Florence (n=132) gathered under the auspices of the DRAGON
352 consortium. A summary of the patient characteristics is described in Supplementary
353 Table 1. Within this cohort ten patients had fatal disease. As no outcome variables
354 within this cohort had power, an unsupervised approach was undertaken. Out of the
355 173 patients, 109 patients had matched CT data scored by clinicians. The data is
356 summarised in Supplementary Table 2. The majority of patients had a CORADS score
357 of high and very high, where 26% was equivocal, 4.6% low and 2.8% very low. The
358 CORADS score stands for "COVID-19 Reporting and Data System," which is a
359 classification system used in radiology to assess the likelihood of COVID-19 infection
360 based on chest imaging findings, typically on computed tomography (CT) scans. The
361 score categorizes imaging findings into different levels of suspicion for COVID-19,
362 ranging from very low to very high.

363

364 *Table 1: Characteristics of 132 patients from Florence included in the study,*
365 *including lab results at admission.*

Characteristic	N	N = 132[1]
Died	132	
N		127 (96%)
Y		5 (3.8%)
Age	132	60 (50, 68)
Sex	132	
F		40 (30%)
M		92 (70%)
Non-invasive ventilation	132	

N		123 (93%)
Y		9 (6.8%)
Continuous positive airway pressure	132	
N		129 (98%)
Y		3 (2.3%)
Tracheostomy	132	
N		131 (99%)
Y		1 (0.8%)
High flow nasal cannula oxygen therapy	132	
N		105 (80%)
Y		27 (20%)
Hypertension	132	
N		77 (58%)
Y		55 (42%)
Malnutrition	132	
N		131 (99%)
Y		1 (0.8%)
Cardiovascular disease	132	
N		119 (90%)
Y		13 (9.8%)
Respiratory disease	132	
N		118 (89%)
Y		14 (11%)

Cancer	132	
N		118 (89%)
Y		14 (11%)
Chronic kidney disease	132	
N		130 (98%)
Y		2 (1.5%)
Chronic hepatitis	132	
N		130 (98%)
Y		2 (1.5%)
Cerebrovascular disease	132	
N		125 (95%)
Y		7 (5.3%)
Chronic hematologic disease	132	
N		129 (98%)
Y		3 (2.3%)
Diastolic blood pressure (mmHg)	132	79 (70, 85)
Heart rate (BPM)	130	80 (75, 89)
Systolic blood pressure (mmHg)	132	125 (115, 140)
Temperature (°C)	131	36.50 (36.00, 37.20)
Weight (kg)	128	78 (70, 89)
Height (cm)	126	170 (165, 175)
Alanine aminotransferase (U/L)	128	27 (17, 39)
Aspartate aminotransferase (U/L)	64	31 (24, 46)

Bilirubin (mg/dL)	128	0.50 (0.30, 0.60)
Calcium (mg/dL)	96	4.50 (4.34, 4.63)
Creatinine (mg/dL)	130	0.83 (0.73, 0.95)
D-dimer (ng/mL)	91	728 (429, 1,091)
Direct bilirubin (mg/dL)	44	0.25 (0.17, 0.29)
Fibrinogen (mg/dL)	117	572 (446, 654)
Fraction of inspired oxygen (%)	127	28 (21, 36)
Hematocrit (%)	132	42.7 (39.7, 45.8)
Lactate dehydrogenase (U/L)	118	297 (247, 359)
Lactic acid (mg/dL)	102	9.0 (7.0, 11.9)
Leukocytes (10^9/L)	132	6.2 (4.6, 7.7)
Lymphocytes (10^9/L)	129	0.90 (0.68, 1.25)
Neutrophils (10^9/L)	129	4.67 (3.08, 6.16)
Oxygen saturation (%)	109	96.10 (94.20, 97.70)
Partial pressure oxygen (mmHg)	131	74 (65, 87)
Partial pressure carbon dioxide (mmHg)	127	36.2 (34.0, 39.0)
Platelets (10^9/L)	132	196 (156, 255)
Potassium (mmol/L)	128	3.85 (3.50, 4.10)
Procalcitonin (ug/L)	126	0.09 (0.06, 0.15)
Prothrombin time (seconds)	127	13.00 (12.30, 13.70)
Sodium (mmol/L)	129	137 (135, 140)
Urea nitrogen (g/L)	64	30 (30, 50)

[1] **n (%); Median (IQR)**

367 **Patients form 3 clusters based on their pathway activation**
368

369 The RNA sequencing data was used to derive gene expression data (mRNA
370 identification and abundance) which was calculated using Salmon inferred with
371 Tximport in R, where values were converted into log2 counts per million (cpm). TopMD
372 was then employed to calculate the activation of pathways. To identify differences in
373 pathway activation across the cohort, activation data was plotted on a PCA which
374 revealed three distinct clusters of patients (Figure 1). The relationship between clinical
375 observations, demographics and CT scan data in each cluster was explored, and the
376 significant differences are reported in Table 3. Lactic acid was slightly higher in cluster
377 1 and 2 and lower in cluster 3. A higher proportion of respiratory disease was observed
378 in cluster 2 and the fraction of inspired oxygen was also higher in this cluster. Direct
379 bilirubin was also higher in cluster 2. The majority of those that died from COVID-19
380 were in cluster 2. CORADS scoring was unable to distinguish between the clusters at
381 a molecular level.

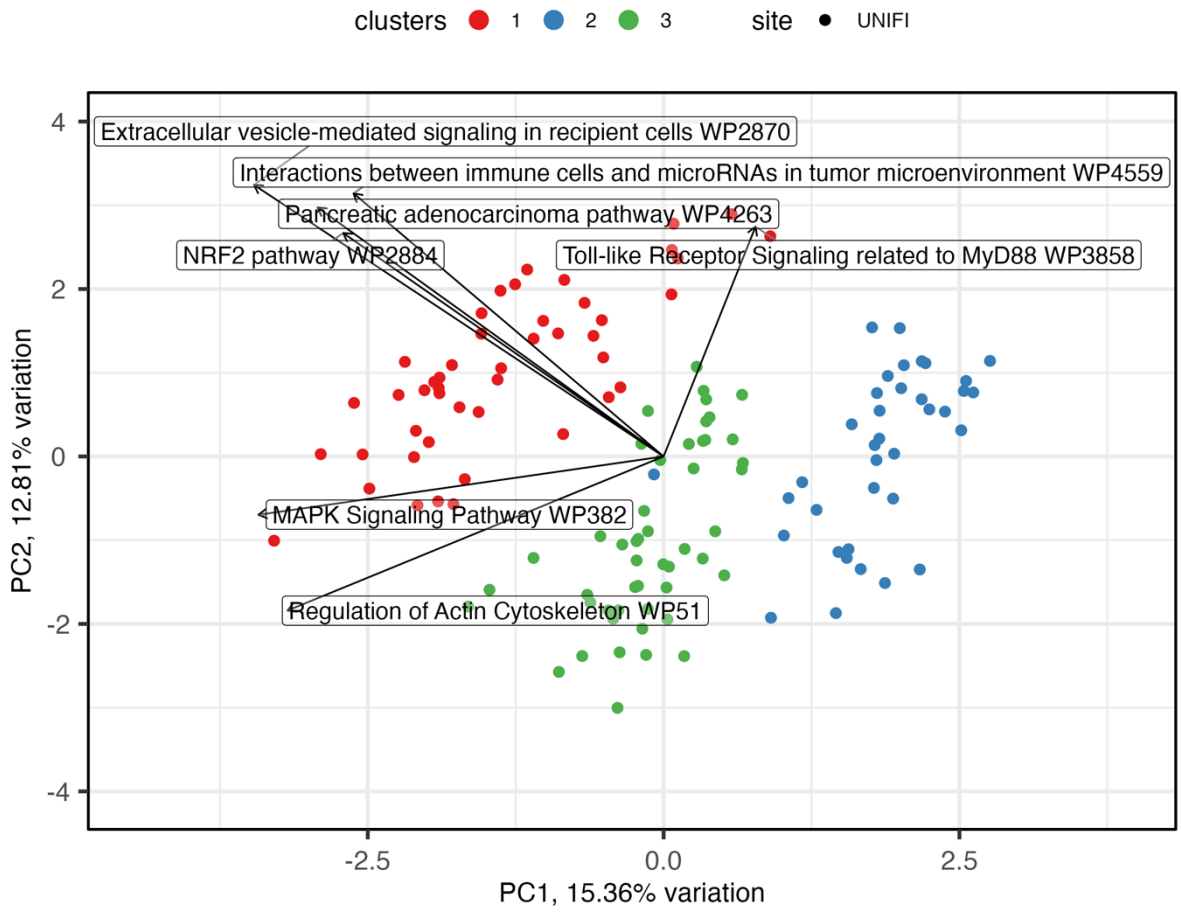


Figure 1:TopMD pathway volumes of each patient in the Florence cohort, calculated from a healthy plotted as a PCA plot. The data reveals three distinct clusters based on pathway activation determined by kmeans.

386 *Table 2: Patient characteristics that differ between the three clusters in the Florence*
387 *cohort (p = <0.05).*

Characteristic	N	**1**, N = 46[1]	**2**, N = 37[1]	**3**, N = 49[1]	**p-value**[2]
Lactic acid (mg/dL)	102	10.0 (7.7, 13.0)	10.0 (7.2, 12.0)	8.0 (5.3, 9.7)	0.008
Fraction of inspired oxygen (%)	127	28 (21, 36)	32 (27, 40)	28 (21, 29)	0.019
Died	132				0.032
N		46 (100%)	33 (89%)	48 (98%)	
Y		0 (0%)	4 (11%)	1 (2.0%)	
Respiratory disease	132				0.042
N		44 (96%)	29 (78%)	45 (92%)	
Y		2 (4.3%)	8 (22%)	4 (8.2%)	
Direct bilirubin (mg/dL)	44	0.20 (0.17, 0.27)	0.28 (0.24, 0.32)	0.20 (0.17, 0.28)	0.047

[1] n (%); Median (IQR)

[2] Fisher's exact test; Kruskal-Wallis rank sum test; Pearson's Chi-squared test

388

389 Molecular phenotype, Cluster 1, was characterised by high activation of pathways
390 associated with ESC pluripotency, NRF2, and TGF-β receptor signalling (Figure 2).
391 Molecular phenotype, Cluster 2 displayed high activation of pathways including focal
392 adhesion-PI3K-Akt-mTOR signalling and type I interferon induction and signalling,
393 while Cluster 3 exhibited low IRF7-related pathway activation.

394

LRROC analysis was conducted on models trained using 70% of patients from the Florence cohort, with test results evaluated on the remaining 30% of the Florence cohort. The area under the ROC curve (AUCROC) values were found to be 0.84, 0.85, and 0.72 for clusters Cluster 1, Cluster 2, and Cluster 3, respectively. Subsequently, these clusters were validated in the Liège cohort (Supplementary Figure 1), yielding AUCROC values of 0.76, 0.93, and 0.69 for Cluster 1, Cluster 2, and Cluster 3, respectively (Supplementary Figure 2).

Potential drug candidates are identified for each cluster

To identify potential drug candidates that modulate pathways identified in these patient clusters, TopMD pathway activation was mapped onto the Drug-Gene Interaction Database (Figure 3). This mapping revealed distinct drug targets for each cluster, detailed in the supplementary table 4. This approach has a two-fold benefit: informing potential clinical trials and informing underlying biological mechanisms specific to each cluster. Interestingly, the pattern of pathway activation might also provide insight into the potential benefits or drawbacks of specific therapies, considering a drug's mechanism of action.

While all clusters shared targetable pathways led by genes such as ITGB2, GNAS, and CXCR2, unique targets also emerged. Cluster 1 specifically identified IFNAR1, TGFBR2, and CSF2RB, while cluster 2 added SERPING1 and TLN1. Notably, cluster 3 shared SERPING1 with cluster 2. These findings highlight both commonalities and variations in potential therapeutic targets across the identified patient clusters.

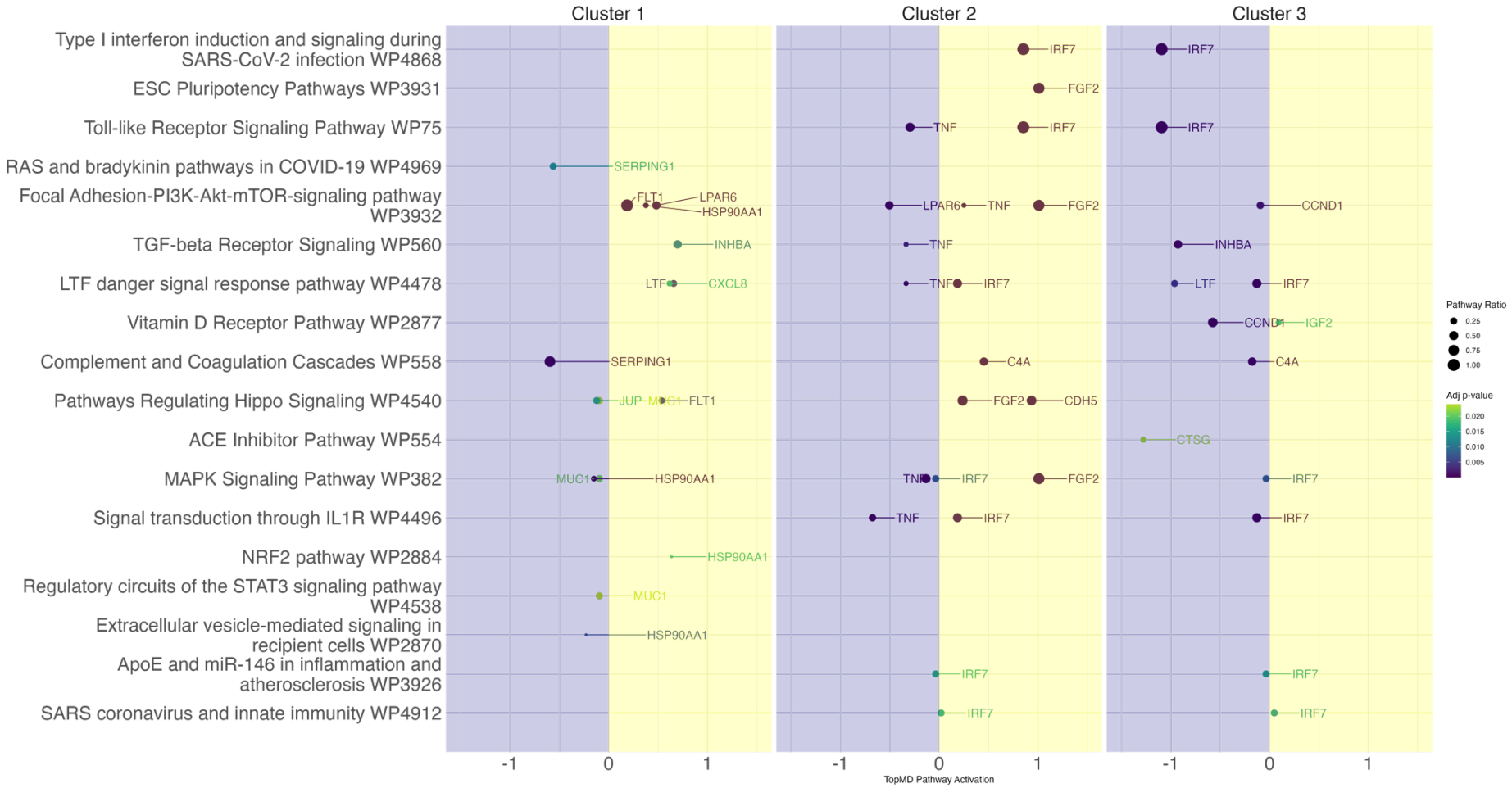


Figure 2: The average pathway volume for each cluster was considered in a TopMD enrichment analysis against the average pathway activation for the whole cohort to identify differentially activated pathways. The enrichment analysis was filtered by adjusted P value, then the top pathways were plotted. The pathways are annotated with the gene that leads the identified pathway. The dots are coloured by adjusted p-value and the size represents the proportion of genes identified within that pathway from TopMD analysis.

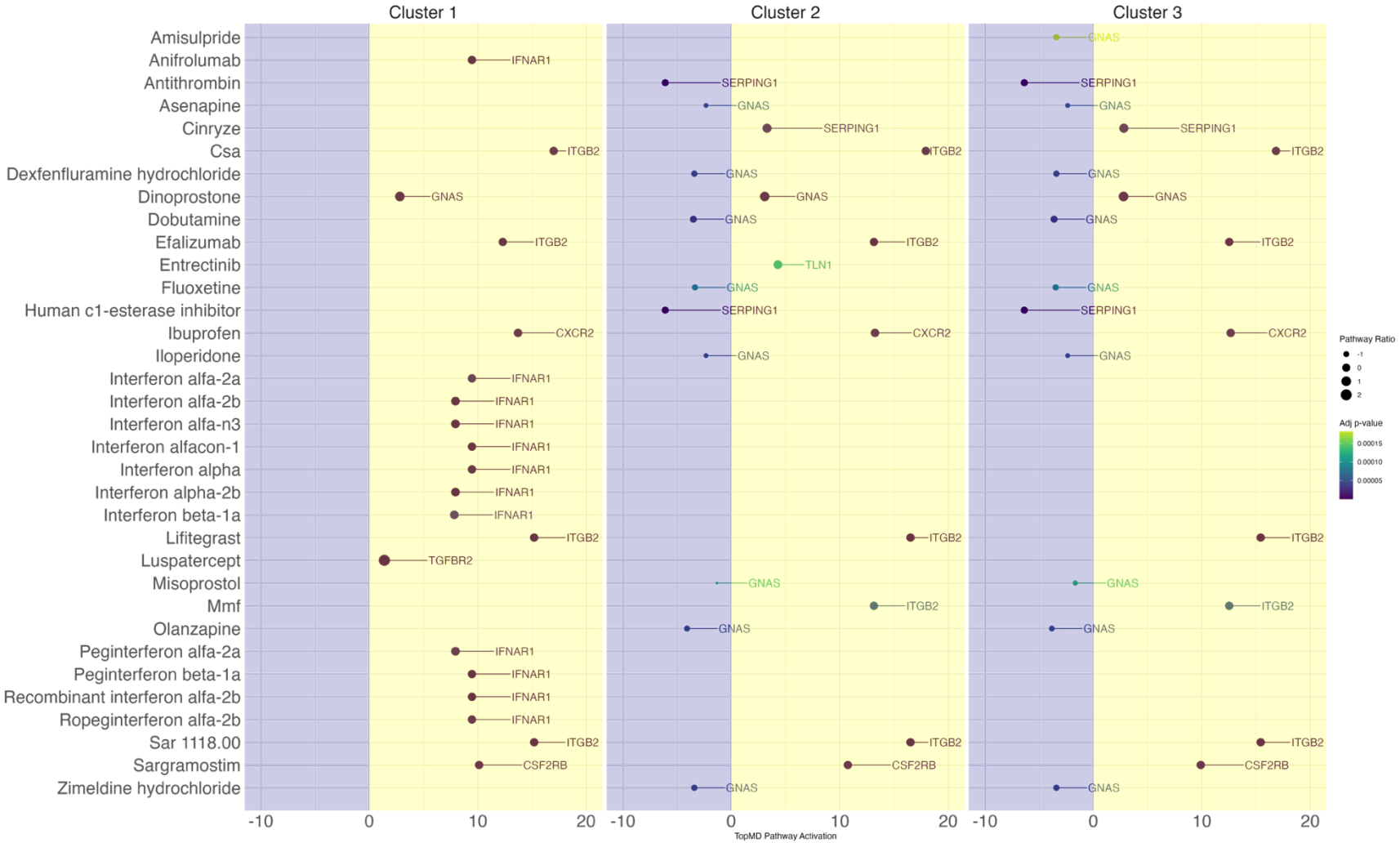


Figure 3: TopMD enrichment analysis was mapped against the Drug-Gene Interaction Database, using a healthy baseline, revealing approved drugs that are known to target genes and their corresponding pathways. The top drug candidates are plotted based on adjusted p-value and pathway volume.

Identification of pathways in fatal cases where intervention might promote survival

Due to limited sample size, we focused on the unsupervised analysis; however, to show utility of investigating pathway activity in individuals, pathway analysis in the 10 deceased patients from the Florence and Liège cohort were observed. Unsurprisingly, these patients exhibited advanced age and high comorbidity rates (cardiovascular: 70%, respiratory: 50%, malnutrition: 40%, hypertension: 90%, cerebrovascular: 30%, chronic hepatitis: 40%). Interestingly, all 10 patients displayed a strong signal for "nonalcoholic fatty liver disease" driven by the NDUFA9 and UQCRC2 genes (Figure 4).

Despite this shared pathway, individual analysis revealed heterogeneity among deceased patients, highlighting the complex interplay between COVID-19, comorbidities, and individual demographics on pathway activation.

Enrichment analysis identified potential therapeutic targets based on individual pathway activation. All patients displayed potential targets including CXCR2 (Figure 5). Additionally, specific druggable pathways were identified for some patients, including GNAS (multiple patients), ITGB2 (patients 2 & 6), CSF2RB (multiple patients), SERPING1 (5 patients), PIK3CD (patient 5), TGFBR2 (patient 9), and CUL4B (patient 10).

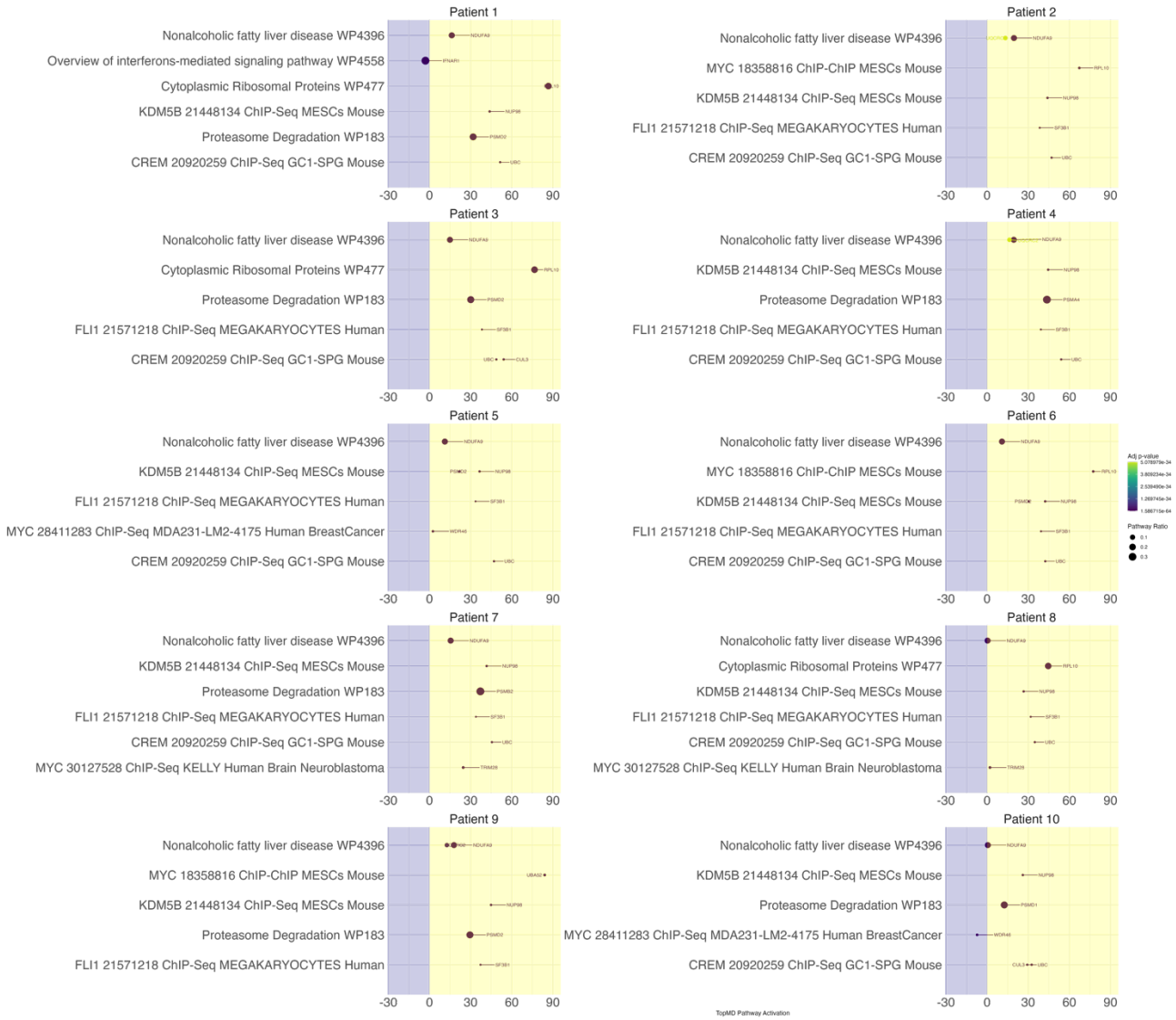


Figure 4: The top 6 pathways enriched in fatal cases within the Florence and Liège cohort using a healthy baseline.

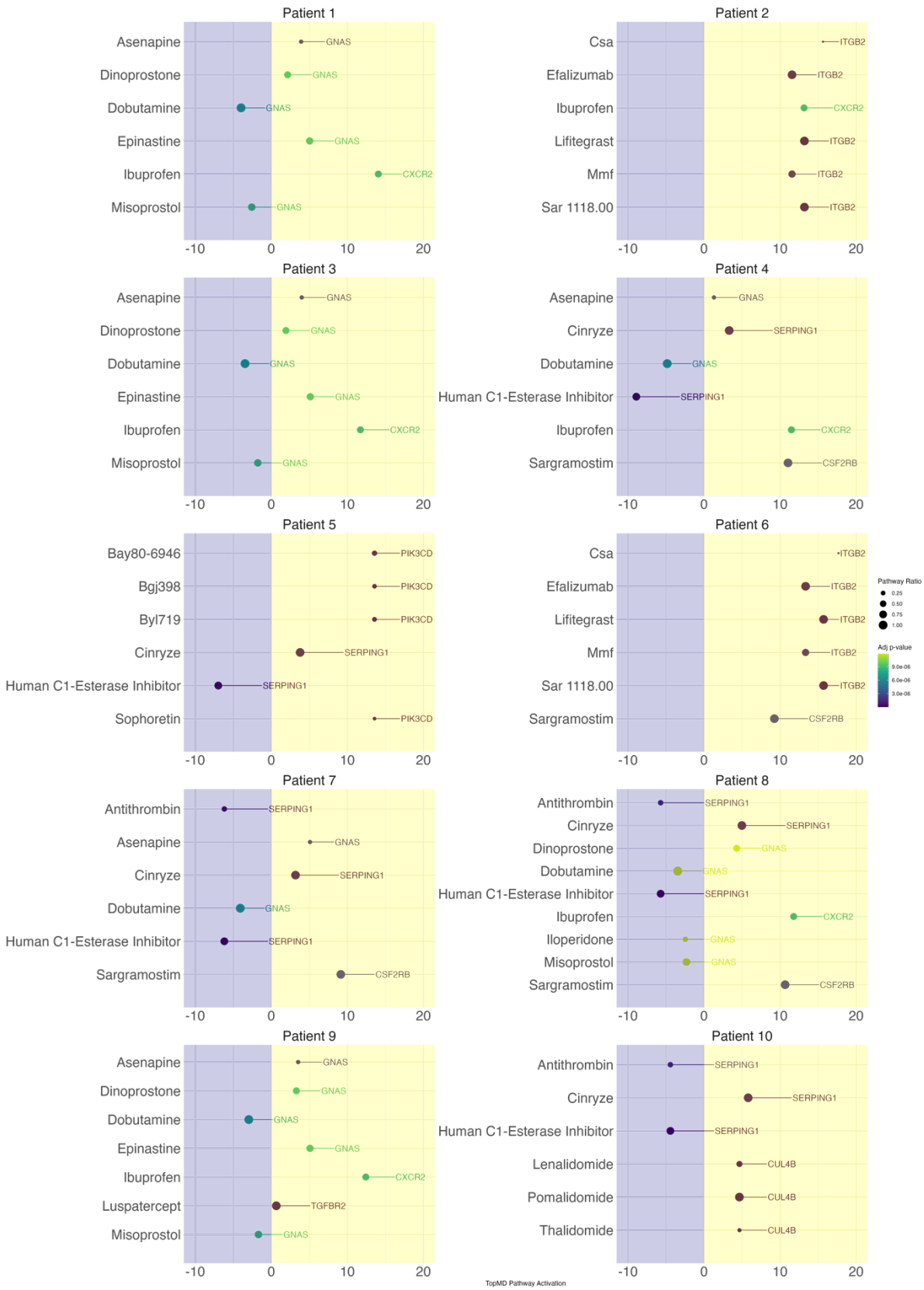


Figure 5: The top significant drug candidates and peak genes that could potentially modulate the phenotype of the 10 fatal cases patients in the Florence and Liège cohort.

468 **Discussion:**

469 Traditionally, molecular phenotyping requires data reduction and feature selection,
470 removing biological and technical 'noise', prior to pathway enrichment analysis, but
471 this leads to results which do not accurately represent the molecular phenotypes.
472 Topological analysis of global gene expression finds value in the low abundance
473 transcripts usually discarded as noise, as they represent the 'foothills' of largely
474 activated pathways in a comprehensive molecular landscape. By understanding the
475 molecular phenotype, it is possible to achieve more successful selection of
476 therapeutics, as medicines work at the molecular level as opposed to a clinical level
477 (28).

478

479 To redefine predictive models for patient outcomes and health trajectories, there is a
480 growing recognition of the importance of integrating complex datasets. This ranges
481 from biomarkers, clinical parameters to CT scans. For instance, a fully automated AI
482 framework was developed to extract features from chest CT scans for diagnosing
483 COVID-19. The model achieved 85.18% accuracy, enabling rapid and accurate
484 differentiation of COVID-19 from routine clinical conditions, facilitating timely
485 interventions and isolation procedures (29). Similarly, an AI-based analysis named
486 CACOVID-CT was implemented to automatically assess disease severity on chest CT
487 scans. Retrospective analysis of 476 patients revealed that quantitative
488 measurements, such as the percentage of affected lung area (% AA) and CT severity
489 score (CT-SS), correlated strongly with hospital length of stay, ICU admission,
490 mechanical ventilation, and in-hospital mortality. This tool proved effective in
491 identifying patients at higher risk of severe outcomes, facilitating patient management
492 and relieving the workload of radiologists (30).

493

494 Our study identified three distinct molecular phenotypes of COVID-19 molecular
495 through topological analysis of global blood gene expression. LRROC analysis
496 demonstrated strong discriminative power of the defined patient clusters tested in the
497 Florence and validated in the Liège cohort. This revealed insights into underlying
498 disease mechanisms, potentially guiding personalised therapeutic approaches.

499

500 The analysis using the TopMD algorithm assigned patients to three clusters. Some of
501 the clinical observations aligned with the defined clusters, including lactic acid

502 elevation in cluster 1 and 2 compared to cluster 3. Elevated lactic acid is known to be
503 associated with disease severity and mortality (31). Similarly, cluster 2 showed a
504 higher proportion of respiratory disease and required a higher fraction of inspired
505 oxygen. Additionally, this cluster exhibited elevated direct bilirubin, another potential
506 indicator of disease severity (32). Notably, the majority of those that died from COVID-
507 19 were in cluster 2 (n=4), although the overall number of fatalities in this cohort was
508 small (n=5).
509
510 Interestingly, the CORADS scoring system used for chest X-ray/CT severity
511 assessment, couldn't differentiate between the molecular clusters. This suggests
512 different molecular mechanisms might underlie similar clinical presentations, which
513 cannot be identified by CT scan. However, utilising higher resolution CT scan data,
514 such as continuous scoring systems offered by tools like Thirona, might provide more
515 granular insights compared to the categorical data used in this study (30).
516
517 Molecular differences were examined between each cluster by considering statistically
518 significant GSEA pathways with highest TopMD pathway volumes (Fig. 2). Cluster 1
519 displayed a reduction in pathways related to the renin-angiotensin system (RAS) and
520 bradykinin, implicated in COVID-19 pathogenesis (33). Additionally, an increase in
521 focal adhesion pathways, possibly indicating cellular changes related to tissue repair
522 and remodelling. Activation of the complement cascade, led by SERPING1, indicates
523 involvement in the immune response to the virus. Furthermore, an increase in the
524 TGF-β pathway, which regulates inflammation and tissue repair was also identified.
525 Additionally, a high activation of pathways associated with ESC pluripotency, NRF2,
526 and TGF-β receptor signalling. The ESC pluripotency pathway is implicated in tissue
527 repair and regeneration, suggesting a potential compensatory response to tissue
528 damage caused by the virus. NRF2 pathway activation may indicate an antioxidant
529 response to counteract oxidative stress induced by viral infection (34). TGF-β receptor
530 signalling, known for its role in regulating inflammation and fibrosis, may contribute to
531 tissue remodelling and fibrosis observed in severe COVID-19 cases (35). Also, cluster
532 1 exhibits low activation of pathways related to extracellular vesicle-mediated
533 signalling and complement and coagulation cascades. The decrease in extracellular
534 vesicle-mediated signalling may reflect impaired intercellular communication, while the

535 low activation of complement and coagulation cascades suggests a possible
536 dysregulated immune response and coagulopathy (36).
537
538 In cluster 2, high activation of pathways such as focal adhesion-PI3K-Akt-mTOR
539 signalling and type I interferon induction and signalling was observed, and has been
540 proposed as a potential therapeutic target in SARS-CoV-2 (37, 38) and MERS-CoV
541 (39). Focal adhesion pathway activation may indicate cellular responses to tissue
542 injury or viral invasion, while type I interferon induction and signalling reflect a strong
543 antiviral immune response (40). In contrast, cluster 3 shows opposite activity in IRF7-
544 related pathways compared to 2. Additionally, vitamin D receptor activity was
545 observed, which has been implicated in modulating the immune response and may
546 play a role in COVID-19 severity (41-43). Notably, this cluster exhibited low activation
547 of pathways related to TGF-β receptor signalling, IL1R signalling, and LTF danger
548 signal response. The reduced TGF-β receptor signalling suggests decreased fibrotic
549 response and tissue remodelling, while low IL1R signalling may indicate attenuated
550 inflammation (44). The activation of the LTF danger signal response pathway appears
551 to be diminished. Lactoferrin demonstrates antiviral capabilities against various
552 viruses, including coronaviruses (45). It can impede viral replication, disrupt viral
553 attachment and entry, and adjust host immune responses. Lactoferrin's
554 immunomodulatory attributes might aid in tempering excessive inflammation and
555 alleviating cytokine storms observed in severe cases of COVID-19 (46). The
556 decreased activation of the LTF danger signal response pathway could potentially
557 contribute to a weakened interferon response (47).
558
559 The stratified molecular phenotypes were found to have different expected responses
560 to both medicines used, and medicines not yet used for COVID-19 (Fig. 3). In cluster
561 1, CSA or cyclosporine has been shown to be safe to use during COVID-19 for the
562 intended use, however, a reduction in hyperinflammation was observed (48). This
563 warrants further investigation as highlighted by others (49). Interferon related therapies
564 that could modulate the pathway activation of cluster 1 were also identified, which have
565 been shown to have positive effects (50-52). Lifitegrast inhibits SARS-CoV-2 *in vitro*
566 (53, 54) By inhibiting TGF-β signalling, Luspatercept may help mitigate the excessive
567 inflammatory response and tissue damage seen in severe COVID-19 cases. Similarly,

568 Sargramostim has shown promise in a small study, but larger trials are needed to
569 confirm these findings (55).
570
571 Like, cluster 1, CSA was also identified as a potentially effective treatment for clusters
572 2 & 3. The mechanisms of actions of other medicines only matched the molecular
573 phenotype of cluster 2. Asenapine, an anti-pyschotic drug identified by others as a
574 potential drug candidate for COVID-19 (56, 57) Cinryze a human c1 esterase inhibitor
575 was also identified, these inhibitors have been shown to improve lung computed
576 tomography scores and increase blood eosinophils, which are indicators of disease
577 recovery, however, time to clinical improvement was not observed (58). Also, for
578 cluster 2, we identified Fluoxetine and other SSRIs such as fluvoxamine which has
579 previously been identified as having potential use for the treatment of COVID-19 and
580 long-COVID (59) Amisulpride was also identified in cluster 3.
581
582 To further evaluate the utility of the TopMD algorithm for precision medicine,
583 enrichment analysis was performed on individual data from the 10 fatal cases within
584 the Florence and Liège cohorts. This approach highlights pathway activation specific
585 to each patient, bypassing the need for a whole cohort for deconvolution. All 10
586 patients showed potential therapeutic targets based on pathway enrichment. CXCR2
587 and GNAS were commonly activated across patients (Figure 5), suggesting drugs
588 such as Ibuprofen may be able to modulate some pathways associated with their
589 phenotype. For patients 2 and 6, ITGB2 emerged as one of the top druggable
590 pathways. Notably, Lifitegrast has shown to inhibit SARS-CoV-2 *in vitro* (53, 54).
591 Additionally, CSA or cyclosporine, was also identified, which was another compound
592 identified in the cluster analysis.
593
594 Multiple patients exhibited CSF2RB enrichment, indicating potential for Sargramostim,
595 a drug shown to reduce mortality and incubation in small COVID-19 study (55).
596 SERPING1 enrichment in 5 patients suggests various approved drugs for pathway
597 modulation, including antithrombin, human c1 esterase inhibitor and cinryze. Patient
598 specific findings were also observed. PIK3CD enrichment in patient 5 suggests
599 Sophoretin as a potentially modulator, with a meta-analysis showing quercetins
600 (including sophoretin), reduce LDH, hospitalisation risk and mortality (60). Patient 9
601 displayed TGFBR2 enrichment indicating luspatercept as a potential drug (identified

602 in the cluster analysis) as a potential candidate. Lastly, CUL4B enrichment in patient
603 10, suggests Thalidomide, Pomalidomide, Lenalidomide for pathway modulation.
604 While Lenalidomide, used to manage multiple myelomas, has been proposed as
605 protective against sever COVID-19 in a case report (61) a clinical trial showed no
606 benefit (62). Thalidomide, although showing no benefit itself (62), remains a subject of
607 discussion for its potential use in COVID-19 (63).
608
609 As a proof of concept, TopMD models were integrated into the Comunicare platform
610 (27), a tool developed and configured within the framework of the DRAGON project,
611 aimed at patient empowerment and providing disease management tooling for
612 clinicians and patients. This proof of concept also enables the analysis of clinical data
613 for clinicians in a dedicated dashboard to demonstrate the possibilities of
614 transcriptomics in digital health. As an example, we generated a model that predicts
615 ICU admission based on our previous work (5) as other outcome variables were too
616 low in number. If a clinician has access to transcriptomic data, a csv file can be
617 uploaded to the dashboard and in return activated pathways are returned after running
618 analysis on the TopMD API. While the use of transcriptomics at the bedside is not
619 ready for deployment, we propose that it is a major advance to be able to demonstrate
620 integration of this data into digital health platforms as the growth of precision medicine
621 continues.
622
623 This study identified three distinct molecular phenotypes in hospitalised COVID-19
624 patients, which were not associated with differences in CT scans and clinical
625 observations. However, these molecular phenotypes match the mechanism of action
626 of different medicines, providing the opportunity for biomarker-led stratified medicine.
627 Topological analysis of global gene expression to define a patient's pathway activation
628 map could be useful in future pandemics to aid in treatment decisions before clinical
629 trials can be completed.
630

Funding:

TopMD, the University of Southampton, Imperial College London, CDISC, Comunicare Solutions, University Hospital of Liège (CHU Liège), and the University of Liverpool are members of the DRAGON consortium. The DRAGON project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 101005122. The JU receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. This publication reflects the author's view. Neither IMI nor the European Union, EFPIA or the DRAGON consortium, are responsible for any use that may be made of the information contained therein. JAH, TP, RPR and CH were supported by the US Food and Drug Administration Medical Countermeasures Initiative (no 75F40120C00085) awarded to JAH and work was also supported by the MRC funded MR/Y004205/1 'The G2P2 virology consortium'. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. DB is funded by NIHR and MRC.

Acknowledgments:

The authors wish to thank the study participants and the hospital staff for their participation in this study.

Contributions:

RPR, FS: methodology, RPR, FS: visualisation, RPR, FS, BVE, TK: software, BE: project administration, RPR, FS, JG, MH, AS, EG, AFD, LG, AJP, CN: investigation, RPR, FS, JG, CH, TP, SW: formal analysis, JG, CN: clinical supervision, RPR, FS, AJP, ST, RB, KH, XX, YN, SW: data curation, RPR, FS, EP, PS, JAH, JPS: writing - original draft preparation, RPR, FB, BE, BE, JG, MH, AS, EG, AFD, LG, AJP, CN, ST, RB, KH, CH, TP, RD, TC, DB, SW, XX, YN, SW, SW, GY, PJS, JAH, JPRS: writing - review and editing, JPRS, JAH, PJS, GY, SW: funding acquisition

Declarations:

RPR is an employee at TopMD Precision Medicine Ltd. JPRS is a founding director, CEO, employee, and shareholder in TopMD Precision Medicine Ltd. FS is a founding director, CTO, employee, and shareholder in TopMD Precision Medicine Ltd. PS is a

founding director, employee, and shareholder in TopMD Precision Medicine Ltd. BVE is CEO of Comunicare Solutions. TK is CTO of Comunicare Solutions.

Data availability:

Sequencing reads available under SRA bioproject: PRJNA1085259

References:

1. Clark JJ, Penrice-Randal R, Sharma P, Kipar A, Dong X, Pennington SH, et al. Sequential infection with influenza A virus followed by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) leads to more severe disease and encephalitis in a mouse model of COVID-19. bioRxiv. 2023:2020.10.13.334532.
2. De Neck S, Penrice-Randal R, Clark JJ, Sharma P, Bentley EG, Kirby A, et al. The Stereotypic Response of the Pulmonary Vasculature to Respiratory Viral Infections: Findings in Mouse Models of SARS-CoV-2, Influenza A and Gammaherpesvirus Infections. Viruses [Internet]. 2023; 15(8).
3. Dorward DA, Russell CD, Um IH, Elshani M, Armstrong SD, Penrice-Randal R, et al. Tissue-Specific Immunopathology in Fatal COVID-19. Am J Respir Crit Care Med. 2021;203(2):192-201.
4. Legebeke J, Lord J, Penrice-Randal R, Vallejo AF, Poole S, Brendish NJ, et al. Evaluating the Immune Response in Treatment-Naive Hospitalised Patients With Influenza and COVID-19. Front Immunol. 2022;13:853265.
5. Penrice-Randal R, Dong X, Shapanis AG, Gardner A, Harding N, Legebeke J, et al. Blood gene expression predicts intensive care unit admission in hospitalised patients with COVID-19. Front Immunol. 2022;13:988685.
6. Russell CD, Valanciute A, Gachanja NN, Stephen J, Penrice-Randal R, Armstrong SD, et al. Tissue Proteomic Analysis Identifies Mechanisms and Stages of Immunopathology in Fatal COVID-19. Am J Respir Cell Mol Biol. 2021.
7. Liu X, Speranza E, Muñoz-Fontela C, Haldenby S, Rickett NY, Garcia-Dorival I, et al. Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. Genome Biology. 2017;18(1):4.
8. McClain MT, Constantine FJ, Nicholson BP, Nichols M, Burke TW, Henao R, et al. A blood-based host gene expression assay for early detection of respiratory viral infection: an index-cluster prospective cohort study. Lancet Infect Dis. 2021;21(3):396-404.
9. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. Nature. 2015;524(7563):97-101.
10. Watson RJ, Tree J, Fotheringham SA, Hall Y, Dong X, Steeds K, et al. Dose-Dependent Response to Infection with Ebola Virus in the Ferret Model and Evidence of Viral Evolution in the Eye. Journal of virology. 2021;95(24):e0083321.
11. Russell CD, Valanciute A, Gachanja NN, Stephen J, Penrice-Randal R, Armstrong SD, et al. Tissue proteomic analysis identifies mechanisms and stages of immunopathology in fatal COVID-19. American journal of respiratory cell and molecular biology. 2022;66(2):196-205.
12. Nan Y, Ser JD, Walsh S, Schönlieb C, Roberts M, Selby I, et al. Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions. Information Fusion. 2022;82:99-122.
13. Halilaj I, Chatterjee A, van Wijk Y, Wu G, van Eeckhout B, Oberije C, et al. Covid19Risk.ai: An Open Source Repository and Online Calculator of Prediction Models for Early Diagnosis and Prognosis of Covid-19. BioMed. 2021;1(1):41-9.
14. Chatterjee A, Wu G, Primakov S, Oberije C, Woodruff H, Kubben P, et al. Can predicting COVID-19 mortality in a European cohort using only demographic and comorbidity data surpass age-based prediction: An externally validated study. PLOS ONE. 2021;16(4):e0249920.

15. Cen X, Wang F, Huang X, Jovic D, Dubee F, Yang H, et al. Towards precision medicine: Omics approach for COVID-19. Biosaf Health. 2023;5(2):78-88.
16. Teodori L, Osimani B, Isidoro C, Ramakrishna S. Mass versus personalized medicine against COVID-19 in the “system sciences” era. Cytometry Part A. 2022;101(12):995-9.
17. Wang Z, He Y. Precision omics data integration and analysis with interoperable ontologies and their application for COVID-19 research. Briefings in Functional Genomics. 2021;20(4):235-48.
18. Venkataraman T, Coleman CM, Frieman MB. Overactive Epidermal Growth Factor Receptor Signaling Leads to Increased Fibrosis after Severe Acute Respiratory Syndrome Coronavirus Infection. Journal of virology. 2017;91(12).
19. Venkataraman T, Frieman MB. The role of epidermal growth factor receptor (EGFR) signaling in SARS coronavirus-induced pulmonary fibrosis. Antiviral Res. 2017;143:142-50.
20. Vagapova ER, Lebedev TD, Prassolov VS. Viral fibrotic scoring and drug screen based on MAPK activity uncovers EGFR as a key regulator of COVID-19 fibrosis. Scientific Reports. 2021;11(1):11234.
21. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90.
22. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14(4):417-9.
23. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521.
24. 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.
25. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-d46.
26. Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, et al. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res. 2021;49(D1):D1144-d51.
27. Duquenne JB, Corhay JL, Louis R, Van Cauwenberge H. [Feasibility and effectiveness study of a simplified mobile self-education and self-monitoring application for patients with severe chronic obstructive pulmonary disease]. Rev Med Liege. 2022;77(2):110-7.
28. Farahani M, Niknam Z, Mohammadi Amirabad L, Amiri-Dashatan N, Koushki M, Nemati M, et al. Molecular pathways involved in COVID-19 and potential pathway-based therapeutic targets. Biomedicine & Pharmacotherapy. 2022;145:112420.
29. Guiot J, Vaidyanathan A, Deprez L, Zerka F, Danthine D, Frix A-N, et al. Development and Validation of an Automated Radiomic CT Signature for Detecting COVID-19. Diagnostics. 2021;11(1):41.
30. Guiot J, Maes N, Winandy M, Henket M, Ernst B, Thys M, et al. Automatized lung disease quantification in patients with COVID-19 as a predictive tool to assess hospitalization severity. Frontiers in Medicine. 2022;9.

31. Carpenè G, Onorato D, Nocini R, Fortunato G, Rizk JG, Henry BM, et al. Blood lactate concentration in COVID-19: a systematic literature review. Clin Chem Lab Med. 2022;60(3):332-7.
32. Chen W, Liu H, Yang G, Wang W, Liu Q, Huang C, et al. Effect of Direct Bilirubin Level on Clinical Outcome and Prognoses in Severely/Critically Ill Patients With COVID-19. Front Med (Lausanne). 2022;9:843505.
33. Garvin MR, Alvarez C, Miller JI, Prates ET, Walker AM, Amos BK, et al. A mechanistic model and therapeutic interventions for COVID-19 involving a RAS-mediated bradykinin storm. Elife. 2020;9.
34. Lee C. Therapeutic Modulation of Virus-Induced Oxidative Stress via the Nrf2-Dependent Antioxidative Pathway. Oxid Med Cell Longev. 2018;2018:6208067.
35. Chen J, Wu W, Wang W, Tang Y, Lan H-Y. Role of TGF-β Signaling in Coronavirus Disease 2019. Integrative Medicine in Nephrology and Andrology. 2022;9(1):9.
36. Conway EM, Mackman N, Warren RQ, Wolberg AS, Mosnier LO, Campbell RA, et al. Understanding COVID-19-associated coagulopathy. Nat Rev Immunol. 2022;22(10):639-49.
37. Fattahi S, Khalifehzadeh-Esfahani Z, Mohammad-Rezaei M, Mafi S, Jafarinia M. PI3K/Akt/mTOR pathway: a potential target for anti-SARS-CoV-2 therapy. Immunol Res. 2022;70(3):269-75.
38. Khezri MR, Varzandeh R, Ghasemnejad-Berenji M. The probable role and therapeutic potential of the PI3K/AKT signaling pathway in SARS-CoV-2 induced coagulopathy. Cell Mol Biol Lett. 2022;27(1):6.
39. Kindrachuk J, Ork B, Hart BJ, Mazur S, Holbrook MR, Frieman MB, et al. Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis. Antimicrob Agents Chemother. 2015;59(2):1088-99.
40. Channappanavar R, Perlman S. Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. Semin Immunopathol. 2017;39(5):529-39.
41. Azmi A, Rismani M, Pourmontaseri H, Mirzaii E, Niknia S, Miladpour B. The role of vitamin D receptor and IL-6 in COVID-19. Mol Genet Genomic Med. 2023;11(7):e2172.
42. Hurst EA, Mellanby RJ, Handel I, Griffith DM, Rossi AG, Walsh TS, et al. Vitamin D insufficiency in COVID-19 and influenza A, and critical illness survivors: a cross-sectional study. BMJ Open. 2021;11(10):e055435.
43. Evans RM, Lippman SM. Shining Light on the COVID-19 Pandemic: A Vitamin D Receptor Checkpoint in Defense of Unregulated Wound Healing. Cell Metabolism. 2020;32(5):704-9.
44. Biernacka A, Dobaczewski M, Frangogiannis NG. TGF-β signaling in fibrosis. Growth Factors. 2011;29(5):196-202.
45. Puddu P, Valenti P, Gessani S. Immunomodulatory effects of lactoferrin on antigen presenting cells. Biochimie. 2009;91(1):11-8.
46. Zimecki M, Actor JK, Kruzel ML. The potential for Lactoferrin to reduce SARS-CoV-2 induced cytokine storm. Int Immunopharmacol. 2021;95:107571.
47. Actor JK, Hwang SA, Kruzel ML. Lactoferrin as a natural immune modulator. Curr Pharm Des. 2009;15(17):1956-73.
48. Blumberg EA, Noll JH, Tebas P, Fraietta JA, Frank I, Marshall A, et al. A phase I trial of cyclosporine for hospitalized patients with COVID-19. JCI Insight. 2022;7(11).

49. Devaux CA, Melenotte C, Piercecchi-Marti MD, Delteil C, Raoult D. Cyclosporin A: A Repurposable Drug in the Treatment of COVID-19? Front Med (Lausanne). 2021;8:663708.
50. Kamyshnyi A, Koval H, Kobevko O, Buchynskyi M, Oksenych V, Kainov D, et al. Therapeutic Effectiveness of Interferon-α2b against COVID-19 with Community-Acquired Pneumonia: The Ukrainian Experience. Int J Mol Sci. 2023;24(8).
51. Reis G, Moreira Silva EAS, Medeiros Silva DC, Thabane L, Campos VHS, Ferreira TS, et al. Early Treatment with Pegylated Interferon Lambda for Covid-19. New England Journal of Medicine. 2023;388(6):518-28.
52. Monk PD, Marsden RJ, Tear VJ, Brookes J, Batten TN, Mankowski M, et al. Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: a randomised, double-blind, placebo-controlled, phase 2 trial. The Lancet Respiratory Medicine. 2021;9(2):196-206.
53. Shen X-R, Geng R, Li Q, Chen Y, Li S-F, Wang Q, et al. ACE2-independent infection of T lymphocytes by SARS-CoV-2. Signal Transduction and Targeted Therapy. 2022;7(1):83.
54. Day CJ, Bailly B, Guillon P, Dirr L, Jen FE, Spillings BL, et al. Multidisciplinary Approaches Identify Compounds that Bind to Human ACE2 or SARS-CoV-2 Spike Protein as Candidates to Block SARS-CoV-2-ACE2 Receptor Interactions. mBio. 2021;12(2).
55. Paine R, Chasse R, Halstead ES, Nfonoyim J, Park DJ, Byun T, et al. Inhaled Sargramostim (Recombinant Human Granulocyte-Macrophage Colony-Stimulating Factor) for COVID-19-Associated Acute Hypoxemia: Results of the Phase 2, Randomized, Open-Label Trial (iLeukPulm). Mil Med. 2022;188(7-8):e2629-38.
56. Ku KB, Shin HJ, Kim HS, Kim BT, Kim SJ, Kim C. Repurposing Screens of FDA-Approved Drugs Identify 29 Inhibitors of SARS-CoV-2. J Microbiol Biotechnol. 2020;30(12):1843-53.
57. Rajput A, Thakur A, Rastogi A, Choudhury S, Kumar M. Computational identification of repurposed drugs against viruses causing epidemics and pandemics via drug-target network analysis. Comput Biol Med. 2021;136:104677.
58. Mansour E, Palma AC, Ulaf RG, Ribeiro LC, Bernardes AF, Nunes TA, et al. Safety and Outcomes Associated with the Pharmacological Inhibition of the Kinin-Kallikrein System in Severe COVID-19. Viruses. 2021;13(2).
59. Hashimoto K. Overview of the potential use of fluvoxamine for COVID-19 and long COVID. Discov Ment Health. 2023;3(1):9.
60. Ziaei S, Alimohammadi-Kamalabadi M, Hasani M, Malekahmadi M, Persad E, Heshmati J. The effect of quercetin supplementation on clinical outcomes in COVID-19 patients: A systematic review and meta-analysis. Food Science & Nutrition. 2023;11(12):7504-14.
61. Al Sbihi A, Manasrah N, Sano D. Can Lenalidomide Protect against Severe COVID-19 Symptoms in Multiple Myeloma Patients? A Case Series and Review of the Literature. Eur J Case Rep Intern Med. 2022;9(3):003216.
62. Amra B, Ashrafi F, Torki M, Hashemi M, Shirzadi M, Soltaninejad F, et al. Thalidomide for the Treatment of COVID-19 Pneumonia: A Randomized Controlled Clinical Trial. Adv Biomed Res. 2023;12:14.
63. Sundaresan L, Giri S, Singh H, Chatterjee S. Repurposing of thalidomide and its derivatives for the treatment of SARS-coV-2 infections: Hints on molecular action. Br J Clin Pharmacol. 2021;87(10):3835-50.