Biomedicine & Pharmacotherapy 154 (2022) 113617

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

journal homepage: www.elsevier.com/locate/biopha

Genome-wide transcriptional profiling of pulmonary functional sequelae in ARDS- secondary to SARS-CoV-2 infection☆

María C. García-Hidalgo [a,b,1], Rafael Peláez [c,1], Jessica González [a,b], Sally Santisteve [a], Iván D. Benítez [a,b], Marta Molinero [a,b], Manel Perez-Pons [a,b], Thalía Belmonte [a,b], Gerard Torres [a,b], Anna Moncusí-Moix [a,b], Clara Gort-Paniello [a,b], Maria Aguilà [a], Faty Seck [a,b], Paola Carmona [a], Jesús Caballero [d], Carme Barberà [e], Adrián Ceccato [b,f], Laia Fernández-Barat [b,g], Ricard Ferrer [b,h], Dario Garcia-Gasulla [i], Jose Ángel Lorente-Balanza [b,j], Rosario Menéndez [b,k], Ana Motos [b,g], Oscar Peñuelas [b,j], Jordi Riera [b,h], Jesús F. Bermejo-Martin [b,l], Antoni Torres [b,m], Ferran Barbé [a,b], David de Gonzalo-Calvo [a,b,*,1], Ignacio M. Larráyoz [c,n,**,1]

[a] Translational Research in Respiratory Medicine, University Hospital Arnau de Vilanova and Santa Maria, IRBLleida, Lleida, Spain
[b] CIBER of Respiratory Diseases (CIBERES), Institute of Health Carlos III, Madrid, Spain
[c] Biomarkers and Molecular Signaling Group, Neurodegenerative Diseases Area Center for Biomedical Research of La Rioja, CIBIR, Logroño, Spain
[d] Grup de Recerca Medicina Intensiva, Intensive Care Department Hospital Universitari Arnau de Vilanova, Lleida, Spain
[e] Intensive Care Department, University Hospital Santa María, IRBLleida, Lleida, Spain
[f] Hospital Universitari Sagrat Cor, Barcelona, Spain
[g] Servei de Pneumologia, Hospital Clinic; Universitat de Barcelona; IDIBAPS, Barcelona, Spain
[h] Intensive Care Department, Vall d'Hebron Hospital Universitari. SODIR Research Group, Vall d'Hebron Institut de Recerca (VHIR), Spain
[i] Barcelona Supercomputing Center (BSC), Barcelona, Spain
[j] Hospital Universitario de Getafe, Madrid, Spain
[k] Pulmonology Service, University and Polytechnic Hospital La Fe, Valencia, Spain
[l] Hospital Universitario Río Hortega de Valladolid, Valladolid, Spain; Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, Spain
[m] Pneumology Department, Clinic Institute of Thorax (ICT), Hospital Clinic of Barcelona, Insitut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), ICREA, University of Barcelona (UB), Barcelona, Spain
[n] GRUPAC, Department of Nursing, University of La Rioja, Logroño, Spain

ARTICLE INFO

Keywords:
Acute respiratory distress syndrome
Apoptosis
Diffusion impairment
Post-COVID
RNA-seq

ABSTRACT

Background: Up to 80% of patients surviving acute respiratory distress syndrome (ARDS) secondary to SARS-CoV-2 infection present persistent anomalies in pulmonary function after hospital discharge. There is a limited understanding of the mechanistic pathways linked to post-acute pulmonary sequelae.
Aim: To identify the molecular underpinnings associated with severe lung diffusion involvement in survivors of SARS-CoV-2-induced ARDS.
Methods: Survivors attended to a complete pulmonary evaluation 3 months after hospital discharge. RNA sequencing (RNA-seq) was performed using Illumina technology in whole-blood samples from 50 patients with moderate to severe diffusion impairment (D_{LCO}<60%) and age- and sex-matched individuals with mild-normal

Abbreviation: ARDS, acute respiratory distress syndrome; bp, base pair; AUC, area under the receiver operating characteristic curve; D_{LCO}, lung diffusing capacity for carbon monoxide; GTEx, Genotype-Tissue Expression; ICU, intensive care unit; IMV, invasive mechanical ventilation; NK, natural killer; PCA, principal component analysis; RNase, ribonuclease; RNA-seq, RNA sequencing; ROC, receiver operating characteristic.

☆ on behalf of the CIBERESUCICOVID Project (COV20/00110, ISCIII).

* Correspondence to: Translational Research in Respiratory Medicine, University Hospital Arnau de Vilanova and Santa Maria, IRBLleida, Avda. Alcalde Rovira Roure 80, Lleida 25198, Spain.

** Correspondence to: Biomarkers and Molecular Signaling Group, Neurodegenerative Diseases Area, Center for Biomedical Research of La Rioja, CIBIR. C. Piqueras, 98, Logroño 26006, Spain.

E-mail addresses: dgonzalo@irblleida.cat (D. de Gonzalo-Calvo), ilarrayoz@riojasalud.es (I.M. Larráyoz).

[1] Equally contributed.

https://doi.org/10.1016/j.biopha.2022.113617
Received 13 July 2022; Received in revised form 18 August 2022; Accepted 27 August 2022
Available online 30 August 2022
0753-3322/© 2022 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

lung function (D_{LCO}≥60%). A transcriptomic signature for optimal classification was constructed using random forest. Transcriptomic data were analyzed for biological pathway enrichment, cellular deconvolution, cell/tissue-specific gene expression and candidate drugs.
Results: RNA-seq identified 1357 differentially expressed transcripts. A model composed of 14 mRNAs allowed the optimal discrimination of survivors with severe diffusion impairment (AUC=0.979). Hallmarks of lung sequelae involved cell death signaling, cytoskeleton reorganization, cell growth and differentiation and the immune response. Resting natural killer (NK) cells were the most important immune cell subtype for the prediction of severe diffusion impairment. Components of the signature correlated with neutrophil, lymphocyte and monocyte counts. A variable expression profile of the transcripts was observed in lung cell subtypes and bodily tissues. One upregulated gene, *TUBB4A*, constitutes a target for FDA-approved drugs.
Conclusions: This work defines the transcriptional programme associated with post-acute pulmonary sequelae and provides novel insights for targeted interventions and biomarker development.

1. Introduction

Genome-wide transcriptional profiling constitutes a powerful tool to analyze gene expression with a huge potential for identifying novel drug targets, improving current therapies and discovering biomarkers [1,2]. In particular, whole-blood transcriptomic analysis has gained importance in clinical research [3]. It is performed in highly accessible and homogeneous samples, avoiding the bias caused by tissue sampling, provides a comprehensive picture of the host response to pathological insults and offers information on the transcriptional programs activated in bodily tissues [4,5]. Currently, RNA sequencing (RNA-seq) is considered the gold standard among transcriptome profiling tools [6]. Indeed, the implementation of RNA-seq allows to decipher pathological processes underlying multiple conditions, including respiratory disease [7].

Acute respiratory distress syndrome (ARDS) is a life-threatening respiratory complication characterized by an uncontrolled inflammatory process that leads to the damage of the alveolar endothelial-epithelial barrier and, consequently, to the accumulation of protein-rich edema and the migration of immune cells [8,9]. Although the mortality rate has significantly decreased in the last decade, a great percentage of survivors of ARDS are affected by abnormalities in pulmonary structure and function [10]. Reduction of lung diffusion capacity for carbon monoxide (D_{LCO}) is the most common alteration, being described in 33–82% of patients several months after ARDS resolution [11,12]. In survivors of ARDS secondary to SARS-CoV-2 infection, our group reported an abnormal D_{LCO} in up to 80% of the patients at a 3-month follow-up [13]. These results have been corroborated in different cohorts [14,15].

Gaining a better understanding of the molecular underpinnings that mediate persistent lung dysfunction could aid the development of therapeutics and biomarkers for post-COVID sequelae [16]. In this context, whole-blood transcriptomic analysis will reveal unexpected or unknown mediators in the pathogenesis of post-acute lung sequelae. To address this, we used unbiased next-generation sequencing to investigate the transcriptional programs associated with severe pulmonary diffusion abnormalities in survivors of SARS-CoV-2-induced ARDS.

2. Methods

2.1. Ethics statements

The study was approved by the medical ethics committee of Hospital Universitari Arnau de Vilanova (registry number CEIC/2273). The patients received written information about the nature of the study and provided informed consent. The protocol was performed in full compliance with the Declaration of Helsinki. Demographic, clinical, pharmacological and laboratory data were extracted from the electronic medical records and introduced into a REDCap database with restricted access.

2.2. Study population

This is a substudy of the ongoing multicenter study CIBERESUCICOVID, registered at www.clinicaltrials.gov with the identification NCT04457505 [17]. Patients with severe COVID-19 admitted to the Hospital Universitari Arnau de Vilanova-Santa María (Lleida, Spain) were enrolled if they fulfilled the following criteria: aged over 18, developed ARDS according to the Berlin definition during the hospital stay [18] and attended a "Post-COVID" evaluation 3 months after hospital discharge (between June and December 2020, median [P_{25}; P_{75}] of 92.5 [83.2; 106] days) (Supplemental Fig. S1). The standard pulmonary evaluation was performed as previously described by our group (González et al., 2021, Chest). In brief, D_{LCO} was evaluated using a flow spirometer (MasterScreen, Jaeger, Germany) in agreement with the guidelines of the American Thoracic Society (Celli et al., 2004, Eur. Respir. J.). D_{LCO} was expressed as a percentage of the predicted D_{LCO} value according to the European Community Lung Health Survey (Roca et al., 1998, Eur. Respir. J.). Based on the ATS/ERS task force: standardization of lung function testing [19], patients were classified into (i) patients with normal or mild diffusion impairment (D_{LCO}≥60%) and (ii) patients with moderate to severe diffusion impairment (D_{LCO}<60%). Patients with incomplete pulmonary function evaluation were excluded from further analysis. Among 100 survivors with complete pulmonary function evaluation and blood samples available, a subpopulation of 50 patients was selected for the genome-wide transcriptional profiling study. To minimize potential confounding, we restricted our analysis to survivors with D_{LCO}< 60% and D_{LCO}≥ 60% (ratio 1:2) who matched according to age and sex. Matching was performed using the nearest neighbor method, which is based on measuring the distances between the propensity score of each sample and its pair [20].

2.3. Sample collection

Samples were processed in standardized conditions with support by IRBLleida Biobank (B.0000682) and "Plataforma Biobancos PT20/00021". Venous blood samples were collected in Tempus TM blood RNA tubes (Thermo Fisher Scientific, MA, USA) by venipuncture after a night of fasting and before beginning pulmonary evaluation, according to the manufacturer's instructions. Blood samples were stored at − 80 °C for future analysis.

2.4. Next-generation sequencing and transcriptomic analysis

RNA purification and sequencing were performed by experienced researchers blinded to patient data using standardized and automated protocols. Total RNA was extracted from 200 μL of preprocessed whole blood with the Maxwell® RSC simply RNA Blood Kit (Promega, WI, USA) in combination with the Maxwell® RSC Instruments in agreement with the manufacturer's instructions. The nucleic acid content was quantified with a NanoDrop™ 2000/2000c (Thermo Fisher Scientific, MA, USA) by independent staff.

Transcriptome next-generation sequencing was performed following

Illumina's protocols (San Diego, CA, USA), as previously detailed [21,22]. Briefly, the Fragment Analyzer system (Agilent, CA, USA) was used to assess the quality and integrity of total RNA. RNA was isolated from 1 μg of total RNA using Illumina's Ribo-zero plus rRNA depletion kit (Illumina, Madrid, Spain). The RNA was fragmented into approximately 200 base pair (bp) pieces by using divalent cations under elevated temperature. Reverse transcriptase and random primers were used to generate first strand cDNA from cleaved RNA fragments. DNA polymerase I and RNase H were used to obtain the second strand. Double-stranded cDNA fragments were end-repaired by Klenow DNA polymerase and T4 DNA. Then, cDNA was phosphorylated by T4 polynucleotide kinase and ligated to Illumina indexing adapters. Adapter-tagged libraries were amplified by using 15 cycles of PCR with DNA polymerase (Phusion; Finnzymes Reagents, Vantaa, Finland). Validation and quantification were performed using electrophoresis and qPCR. Pools of six indexed libraries were mixed at equimolar ratios to yield a total oligonucleotide mixture concentration of 10 nM. The resulting libraries were sequenced with a HiSeq 1500 platform (Illumina, Madrid, Spain) to generate 2 × 250-bp paired-end reads using three lanes.

Transcriptomic analyses were performed with paired-end raw read data obtained from Illumina sequencing workflows. Raw sequence data in fastq format were processed using a sequential workflow: (i) data were cleaned of adapters and low-quality sequences using TrimGalore v0.6.6 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ accessed date April 1st, 2021); (ii) reads were mapped to the reference genome (GRCh38.p12) by the aligner STAR (https://github.com/alexdobin/STAR); (iii) mapped reads were counted using FeatureCounts (http://subread.sourceforge.net/ accessed date April 1st, 2021); (iv) global distance analysis and normalization of reads were performed through DESeq2 [23]; (v) differential expression analysis were performed with the R package [24]. Zeros were excluded when less than 25% of samples had at least one count.

2.5. Bioinformatic analyses

The Reactome database was used for the analysis of biological processes/molecular pathways [25]. Significant enrichment was considered if the p-value< 0.05. Cellular deconvolution was performed using the CIBERSORTx algorithm [26]. Raw gene counts were normalized by DESeq2 and processed with the CIBERSORTx gene signature LM22 and 100 permutations. Additionally, a selection process based on random forest [27] was used to assess the variable importance of immune cell fractions. Lung cell and tissue expression was obtained using data from the Genotype-Tissue Expression (GTEx) Portal (https://www.gtexportal.org/home/). The Drug-Gene Interaction database (DGIdb v4.2.0) was used to predict interactions between FDA-approved drugs and genes [28].

2.6. Statistical analysis, data mining and feature selection

Descriptive statistics were used to summarize the characteristics of the study population. Data are presented as the median [25th; 75th percentile] or mean (SD) for continuous variables and as frequencies (percentage) for categorical variables. The normality of the data was analyzed using the Kolmogorov–Smirnov test. Continuous variables were compared using the Mann–Whitney U test or Student's t test, and categorical variables were compared using Fisher's exact test. Spearman's rho coefficient was used to assess the correlation between continuous variables. Correlations between continuous and dichotomous variables were calculated using point-biserial correlation. All statistical analyses were performed using R software, version 4.1.2.

The classification model (boosted random forest) associated with severe diffusion impairment (D_{LCO}<60%) was constructed using open-source data mining tools (available from Weka Machine Learning 3, https://www.cs.waikato.ac.nz/ml/weka accessed date April 1st 2021) [29]. First, an attribute evaluation algorithm was used to reduce the gene sets used for classification modeling. That process was performed in two steps: a model entropy minimization score for each attribute and a correlation-based feature selection algorithm. Filtering of useful probes from transcriptomics data was performed by using the 'InfoGainAttributeEval' feature selection method (attribute evaluator based on the reduction of entropy) followed by the "Ranker" search method to sort the results, from the WEKA suite. This process reduced the set of useful probes for classification 100-fold. The resulting set of probes were further filtered: i) to exclude transcripts expressed less than 5 counts; and ii) transcripts not associated with known genes, in order to simplify the model and to increase the reproducibility of the results in other cohorts. The resulting set fulfills the condition of being the minimal attribute set without compromising model classification performance. Model classification performance was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) to measure model classification performance. The AUC was also used as a performance parameter in the data mining step. The classification model was independently tested using 10-fold cross-validation.

3. Results

3.1. Clinical characteristics of the study population

The most relevant demographic and clinical characteristics of the study population are displayed in Table 1. The mean (SD) age was 57.7 (11.2), and females represented 30% of the population. No differences were observed in sociodemographic characteristics or previous comorbidities, including smoking history or previous pulmonary disease. Except for the duration of prone positioning, similar findings were observed for clinical and ventilatory parameters recorded during the hospital stay. The study population showed a mean (SD) D_{LCO} of 63.9 (14.8). Thirty-six percent of patients showed moderate-severe diffusion impairment (D_{LCO}<60%) at 3 months after discharge. The comparison between those patients included in the study and the whole population is shown in Supplemental Table S1. The study population is a representative sample.

3.2. Whole-blood transcriptome profile associated with severe lung diffusion impairment

Overall, 1357 significant differentially expressed genes were found between the study groups (Fig. 1 A & Supplemental Table S2). To explore the biological relevance of the unfractionated blood transcriptome, the Weka suite was used to implement a multivariable classification model. A model including the 14 transcripts that provided the best discriminative power was constructed using the random forest classifier (Fig. 1B). As shown in the heatmap and PCA (Fig. 1C & D), the classification model was able to segregate patients based on the severity of lung diffusion impairment. The model showed optimal discrimination (AUC=0.979) (Fig. 1E). Among the selected genes, *BAG2*, *ETV7*, *SERPINB6*, *TMEM161B* and *TUBB4A* were upregulated in patients with the most severe pulmonary outcomes (Fig. 1F). Decreased expression of *BCL2L11*, *GAS2*, *GRK2*, *PPP1R10* and *TBC1D10B* was observed in patients with severe diffusion impairment (Fig. 1F). Specific components of the classification model correlated (rho>0.3) with indicators of disease severity and/or patient management during the acute phase, i.e., use and duration of prone positioning, ICU stay, invasive mechanical ventilation (IMV) duration and non-IMV duration (Supplemental Fig. S2).

3.3. Bioinformatic analyses

The blood transcriptomic classification model was then considered for pathway analysis using the Reactome database (Fig. 2A). The analysis identified a number of processes implicated in cell death/apoptosis:

Table 1
Characteristics of the study population.

	All	D_{LCO}≥ 60% predicted	D_{LCO}< 60% predicted	p-value	N	
	N = 50	N = 32	N = 18			
Sociodemographic characteristics						
Age (years)	57.7 (11.2)	56.1 (10.1)	60.3 (12.9)	0.235	50	
Female	15 (30.0%)	10 (31.2%)	5 (27.8%)	1.000	50	
BMI (kg/m^2)	31.0 (6.2)	31.8 (6.3)	29.6 (5.9)	0.208	50	
Smoking history				0.692	49	
Former	32 (65.3%)	21 (67.7%)	11 (61.1%)			
Non-smoker	14 (28.6%)	9 (29.0%)	5 (27.8%)			
Current	3 (6.12%)	1 (3.23%)	2 (11.1%)			
Comorbidities						
Hypertension	25 (50.0%)	15 (46.9%)	10 (55.6%)	0.768	50	
Type II Diabetes Mellitus	10 (20.0%)	8 (25.0%)	2 (11.1%)	0.295	50	
Obesity	25 (50.0%)	17 (53.1%)	8 (44.4%)	0.768	50	
Cardiovascular disease	4 (8.00%)	1 (3.12%)	3 (16.7%)	0.127	50	
Chronic pulmonary disease	5 (10.0%)	3 (9.38%)	2 (11.1%)	1.000	50	
Asthma	2 (4.00%)	1 (3.12%)	1 (5.56%)	1.000	50	
Chronic kidney disease	1 (2.00%)	1 (3.12%)	0 (0.00%)	1.000	50	
Chronic liver disease	3 (6.00%)	2 (6.25%)	1 (5.56%)	1.000	50	
Hospital stay						
Hospital stay (days)	18.5 [10.0;36.2]	16.5 [10.0;40.0]	24.5 [13.0;28.0]	0.531	50	
ICU admission	39 (78.0%)	25 (78.1%)	14 (77.8%)	1.000	50	
ICU stay (days)	14 [5.00;26.0]	15.5 [5.00;27.5]	12 [5.50;15.5]	0.340	39	
Worst PaO_2/FiO_2	138 [90.2;194]	140 [93.2;196]	118 [90.2;193]	0.649	50	
PaO_2/FiO_2 categories				0.655		50
PaO_2/FiO_2 201–300 mmHg	12 (24.0%)	8 (25.0%)	4 (22.2%)			
PaO_2/FiO_2 101–200 mmHg	20 (40.0%)	14 (43.8%)	6 (33.3%)			
$PaO_2/FiO_2 \leq$ 100 mmHg	18 (36.0%)	10 (31.2%)	8 (44.4%)			
High-flow nasal cannula	24 (48.0%)	15 (46.9%)	9 (50.0%)	1.000	50	
IMV	25 (50.0%)	16 (50.0%)	9 (50.0%)	1.000	50	
IMV duration (days)	17 [11.0;25.0]	18 [13.0;25.0]	16 [8.00;25.0]	0.386	24	
Non-IMV	30 (60.0%)	19 (59.4%)	11 (61.1%)	1.000	50	
Non-IMV duration (days)	2 [2.00;4.00]	2 [2.00;4.00]	2 [1.50;6.00]	0.842	30	
Prone positioning	24 (48.0%)	14 (43.8%)	10 (55.6%)	0.612	50	
Prone positioning duration (hours)	43.1 (27.7)	54.2 (28.9)	29.7 (20.2)	0.030	22	
Antibiotics	42 (84.0%)	28 (87.5%)	14 (77.8%)	0.436	50	
Hydroxychloroquine	36 (72.0%)	23 (71.9%)	13 (72.2%)	1.000	50	
Tocilizumab	21 (42.0%)	14 (43.8%)	7 (38.9%)	0.971	50	
Corticoids	41 (82.0%)	26 (81.2%)	15 (83.3%)	1.000	50	
Remdesivir	9 (18.0%)	6 (18.8%)	3 (16.7%)	1.000	50	
Interferon beta	4 (10.5%)	3 (12.5%)	1 (7.14%)	1.000	38	
Methylprednisolone	25 (59.5%)	17 (65.4%)	8 (50.0%)	0.507	42	
Dexamethasone	11 (26.2%)	7 (26.9%)	4 (25.0%)	1.000	42	
Lopinavir/ritonavir	35 (70.0%)	23 (71.9%)	12 (66.7%)	0.949	50	
White blood cell counts at "Post-COVID" evaluation						
Leucocyte count ($x10^9/L$)	6.87 (2.04)	6.80 (1.72)	7.00 (2.57)	0.772	50	
Lymphocyte count ($x10^9/L$)	2.25 [1.79;2.97]	2.14 [1.76;2.75]	2.33 [2.02;3.26]	0.218	50	
Neutrophil count ($x10^9/L$)	3.27 [2.58;4.42]	3.47 [2.59;4.29]	3.04 [2.43;4.67]	0.599	50	
Eosinophil count ($x10^9/L$)	0.17 [0.10;0.24]	0.16 [0.10;0.21]	0.18 [0.11;0.26]	0.606	50	
Basophil count ($x10^9/L$)	0.04 [0.03;0.06]	0.04 [0.04;0.06]	0.03 [0.03;0.05]	0.073	50	
Monocyte count ($x10^9/L$)	0.57 (0.17)	0.55 (0.15)	0.61 (0.20)	0.244	50	
Neutrophil-to-lymphocyte ratio	1.42 [1.12;1.94]	1.51 [1.14;2.12]	1.39 [0.99;1.77]	0.266	50	
Pulmonary function at "Post-COVID" evaluation						
D_{LCO}	63.9 (14.8)	72.4 (9.72)	48.9 (9.15)	< 0.001	50	
D_{LCO}				< 0.001	50	
< 60%	18 (36.0%)	0 (0.00%)	18 (100%)			
< 80%	24 (48.0%)	24 (75.0%)	0 (0.00%)			
≥ 80%	8 (16.0%)	8 (25.0%)	0 (0.00%)			

Continuous variables are expressed as mean (SD) or median [P_{25};P_{75}]. Categorical variables are expressed as n (%). BMI: body mass index. D_{LCO}: carbon monoxide diffusing capacity. FiO_2: fraction of inspired oxygen. ICU: intensive care unit. IMV: invasive mechanical ventilation. NLR: neutrophil-to-lymphocyte ratio. PaO_2: oxygen partial pressure.

RUNX3 regulates BCL2L11 (BIM) transcription, FOXO-mediated transcription of cell death genes, Defective intrinsic pathway for apoptosis, Diseases of programmed cell death, BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members, Caspase-mediated cleavage of cytoskeletal proteins, Apoptosis and programmed cell death. A relevant proportion of the processes were also associated with pathways implicated in cytoskeleton processing: Activation of BIM and translocation to the mitochondria, Caspase-mediated cleavage of cytoskeletal proteins, Activation of SMO, Microtubule-dependent trafficking of connexons from the Golgi to the plasma membrane, Transport of connexons to the plasma membrane and Membrane trafficking. In addition, pathways related to cell differentiation, proliferation and survival (FLT3 signaling, FOXO-mediated transcription and Signaling by Hedgehog) and inflammation and regulation of immune function (Transcriptional regulation by RUNX3) were identified.

Next, using the transcriptome data, we estimated the immune cell subtypes and their proportions in whole blood by CIBERSORTx (Fig. 2B & Supplemental Table S3). No significant differences were observed between the study groups. The random forest analysis found that resting natural killer (NK) cells were the most important fraction for the

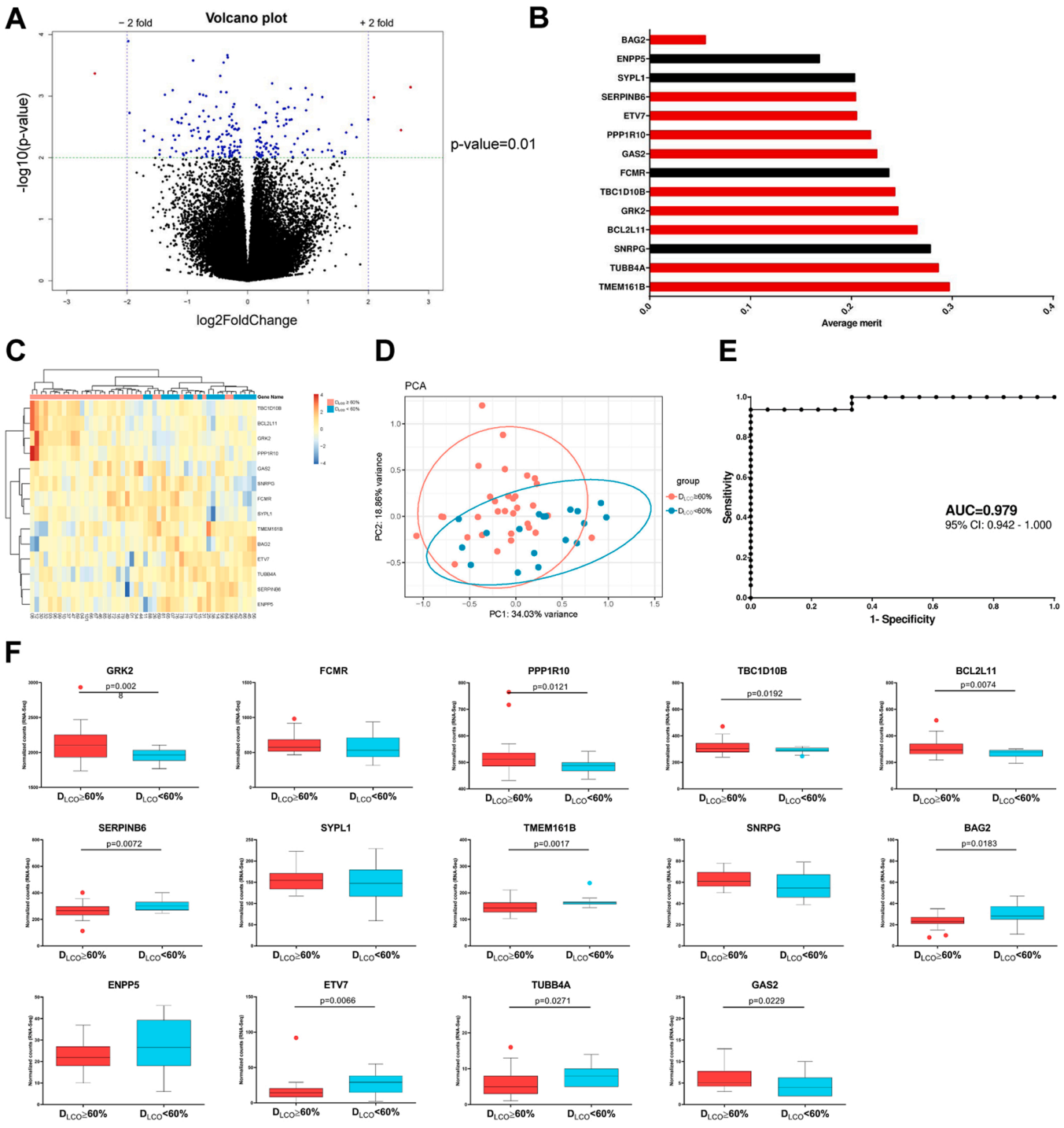


Fig. 1. Transcriptomic signature associated with severe pulmonary diffusion impairment. A) Volcano plot from whole-blood RNA-seq data representing the p-value versus the fold change for each transcript after comparison between D_{LCO}< 60% and D_{LCO}≥ 60%. P-value< 0.01 transcripts are depicted in blue, while FC≥ 2 and p-value< 0.01 genes are represented in red. B) Prediction model based on random forest. Average importance of each selected gene to the model. Black bars indicate non-significantly expressed genes. Red bars reflect significantly expressed genes. C) Heatmap showing unsupervised hierarchical clustering. Each column denotes a patient. Each row denotes one of the 14 genes in the signature. The patient clustering tree is plotted on top. The protein clustering is plotted on the left. Gene levels are represented through a color scale, with red tones associated with increasing levels and blue tones associated with decreasing expression. D) Principal component analysis using the 14 genes in the profile. Each point shows a patient and is represented with a specific color depending on the presence or absence of severe diffusion impairment (D_{LCO}<60%). E) ROC curve for the identified gene signature. The classification performance of the model was assessed using the AUC and 95% confidence interval. F) Boxplots of each of the genes in the identified signature. Comparisons were performed using Welch's corrected or Mann–Whitney t test. The p-value is plotted for each gene.

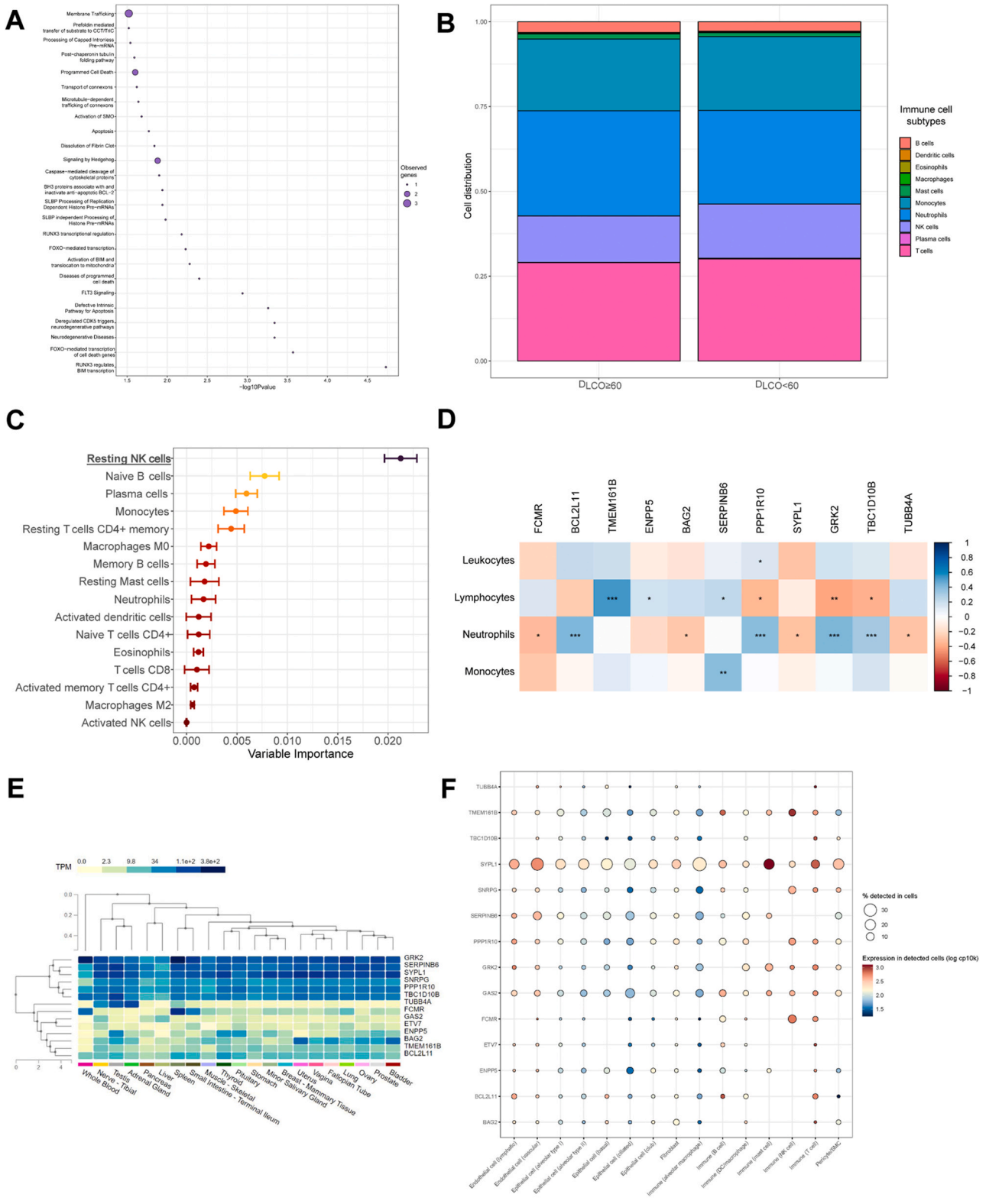


Fig. 2. Pathway enrichment analysis, cellular deconvolution and tissue-/cell-expression profile. A) Gene-set enrichment analysis with gene signature associated with severe pulmonary diffusion impairment using Reactome. The plot shows the p-value of the top twenty-five pathways. The size of each point is proportional to the number of genes of the signature involved in the biological pathway. B) Average immune cell subtype proportion in survivors with D_{LCO}< 60% and D_{LCO}≥ 60%. C) Variable importance plot displaying the most relevant immune cell subtypes according to their contribution to the random forest model. D) Correlation between the gene in the signature and immune cell counts in the follow-up. Significant correlations are marked with asterisks according to the Spearman test. * : p-value< 0.05, ** : p-value< 0.01, *** : p-value< 0.001. E) Tissue enrichment analysis using Genotype-Tissue expression (GTEx). Hierarchical clustering shows tissues at the bottom and genes on the right. F) Cell enrichment analysis based on single-cell RNA-seq for the gene pattern using GTEx Project. Each column shows a cell type, and each row shows a gene. The size of the point shows the number of cells where the gene was detected, and the color represents the expression level. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

prediction of severe diffusion impairment (Fig. 2C). To further explore the association between the transcriptomic signature and immune cell composition, we assessed the correlation between the expression levels of the transcripts in our signature and analytical leukocyte counts (Fig. 2D). Neutrophil counts were directly correlated with *BCL2L11*, *PPP1R10*, *GRK2* and *TBC1D10B* and inversely correlated with *FCMR*, *BAG2*, *SYPL1* and *TUBB4A*. In contrast, lymphocyte counts were directly correlated with *TMEM161B*, *ENPP5* and *SERPINB6* and inversely correlated with *PPP1R10*, *GRK2* and *TBC1D10B*. A positive correlation between monocyte counts and *SERPINB6* was also observed. It has been proposed that the peripheral blood transcriptome shares more than 80% of gene expression with different tissues, including the lung [30]. Using available data from the GTEx project, we analyzed the expression of the genes that composed the model in the different tissues, in addition to lung cells (Fig. 2E & F). The results showed a variable expression pattern in tissues and cells, including epithelial, immune and endothelial lung cells.

A protein-drug interaction analysis was performed using the five upregulated mRNAs. We found six FDA-approved drugs that targeted *TUBB4A* (Supplemental Fig. S3). No interactions were observed for other candidates.

4. Discussion

Here, we use whole blood to comprehensively explore the transcriptomic landscape of post-acute pulmonary sequelae in survivors of COVID-19-driven ARDS. Our analysis revealed a distinct transcriptional program in patients with moderate to severe diffusion impairment 3 months after hospital discharge. Using advanced statistical methods, we constructed a classification model comprised of 14 transcripts that accurately identifies survivors with severe alterations in pulmonary function 3 months after hospital discharge. After implementing integrative pathway enrichment, we identified multiple biological mechanisms implicated in cell death but also in cytoskeleton reorganization, cell survival, growth and differentiation and the regulation of the immune response. To the best of our knowledge, our study is the first to apply blood RNA-seq technology to survivors of ARDS secondary to SARS-CoV-2 infection.

Blood cells play a major role in lung recovery after damage due to infections/pathologic insults [8]. Therefore, an altered immune cell function in survivors of SARS-CoV-2-induced ARDS with severe pulmonary damage is expected. The impaired elimination of harmful/-damaged cells in this group of patients should also be considered. Supporting these hypotheses, our results suggest a profound deregulation of the immune cell phenotype in survivors with moderate to severe abnormalities in lung diffusion capacity. Indeed, severe post-acute sequelae were associated with a complex alteration of cell death signaling, a key mechanism in immune response resolution. In consonance with recently published data from our group [31], enrichment analysis revealed an overrepresentation of apoptotic pathways, including processes regulated and/or mediated by RUNX3 and FOXO and the downstream BIM apoptotic signaling pathway. The individual genes have been described in processes related to immune cell survival and depletion. *BCL2L11* or its coding protein BIM, which is downregulated in survivors with moderate to severe diffusion impairment, is an inducer of apoptosis in activated immune cells [32]. Through the activation of proapoptotic Bax/Bak, BIM is critical for the control of immune system homeostasis, i.e., termination of the immune response and prevention of neoplastic and autoimmune phenomena [33]. In line with this finding, *FCMR* is an inhibitor of the apoptosis regulator Fas that, together with BIM, participates in the shutdown of chronic immune responses and prevention of autoimmunity [34]. BIM has also been described as a lifespan-limiting molecule in immune cells [35]. Cytosolic serpins, such as *SERPINB6*, in addition to preventing an excessive inflammatory response against microbial agents, contribute to myeloid cell survival by modulating programmed necrosis [36]. The upregulation of *BAG2* is compatible with an inhibition of endoplasmic reticulum stress-induced cell apoptosis in macrophages [37] and a promotion in the survival of infected macrophages by inactivating the p53–ROS pathway [38]. Furthermore, *GAS2* is a dual regulator of damage-induced apoptosis in several cell types with a variety of mechanisms of action [39].

Although features of anti-apoptotic signaling seem to be predominant, the impact of the transcriptomic profile on the balance between pro- and anti-apoptotic mechanisms remains to be determined. For example, except for resting NK cells, no clear associations were observed when the fractions of innate and adaptive immunity were inferred. The leukocyte counts support these findings. Nevertheless, the results suggest a close association between moderate to severe lung diffusion impairment and the alteration in immune cell death programs. This may impact the termination of the immune response but may also be an indicator of senescence [40]. Since previous results suggest that immunosenescence could negatively affect tissue healing [41], our results warrant further research. Specific targeting of immunosenescence-related genes through senolytic drugs could be an interesting approach to therapeutic strategies to re-establish homeostasis and facilitate the resolution of damage [42].

Both previous data on individual genes and current analysis suggest a significant enrichment in processes implicated in cytoskeletal architecture. Pathways related to the maintenance and integrity of the cytoskeleton, microtubule-dependent trafficking, activation of transmembrane proteins and transport of cell components were identified. *TUBB4A* is part of the tubulin family, which constitutes the major constituent of microtubules. This gene, upregulated in moderate to severe diffusion impairment and inversely correlated with neutrophil counts, has been previously described as a critical element for the stabilization of microtubule polymerization and cell migration [43]. It is worth noting that the drug–gene interaction analysis also identified several FDA-approved drugs that can target *TUBB4A*. *GAS2* and *GRK2* are important modulators of cytoskeletal remodeling. Both mediators induce a modification of microfilaments structure and subsequently cell membrane ruffling and contraction which, supporting the findings discussed above, are normally coupled to an apoptotic cellular phenotype [44,45]. In this sense, *GAS2* also regulates the interaction between microtubules and actin, which is essential for cell morphology and cytoskeleton reorganization [39,46]. Of note, *GRK2* antagonists have been proposed as a potential immunosuppressant strategy since this gene regulates immune cell migration from blood into tissue [47]. Likewise, *TBC1D10C* is an important modulator of cytoskeletal dynamics in peripheral macrophages and, consequently, might be fundamental for their phagocytosis and spreading capacity [48]. The expression profile could be linked to the infiltration of immune cells from the blood into the lung, which may be relevant for tissue healing [49,50]. Furthermore, the analysis identified relevant genes/pathways implicated in cell survival, growth and differentiation, such as FLT3, FOXO and Hedgehog signaling. In fact, the downregulation of *PPP1R10* and *SNRPG* has been previously linked to reduced cellular proliferation and cell cycle arrest, respectively [51,52]. These results are in accordance with previous data that link proteomic mediators of cell proliferation and differentiation with pulmonary sequelae in severe COVID-19 [16]. Several genes in the classification model, such as *TMEM161B*, *ENPP5*, and *BAG2*, mediate a wide range of cellular processes, including protein synthesis. In particular, *BAG2* is a co-chaperone that promotes Hsp70 chaperone activity and inhibits the degradation of misfolded proteins mediated by the ubiquitin–proteasome system and MAPK [53]. Interestingly, *ETV7*, which is upregulated in survivors with lung diffusion impairment, has been previously described as a negative regulator of the type I IFN response and therefore as a therapeutic target to improve innate antiviral responses and enhance IFN-based antiviral therapies [54]. Whether *ETV7* expression could be a possible indicator of the long-lasting effect of the infection beyond hospitalization deserves further analysis. Overall, functional studies are needed to explore the association between these

mechanistic pathways and post-acute pulmonary sequelae in survivors of SARS-CoV-2-driven ARDS.

COVID-19 still constitutes an important public health challenge due to the significant burden of persistent post-acute sequelae. Survivors of severe clinical courses of the disease suffer a huge decrease in health-related quality of life and organic dysfunction. The use of a significant amount of health care resources should also be highlighted. In this scenario, it is of vital importance to develop tools aimed at assisting in the follow-up of survivors of ARDS derived from COVID-19. Delineating the transcriptional programs of blood cells provides the opportunity to identify novel tools for clinical decision-making. We have shown that survivors with severe pulmonary abnormalities can be distinguished at the transcriptomic level using RNA-seq technology. We present a 14-gene signature that precisely classifies patients according to their pulmonary outcomes 3 months after hospital discharge. Since RNA-based biomarker in vitro assays are commercially available [55], our results are clinically relevant. Profiling blood transcript abundance on a genome-wide scale emerges as a promising tool for the development of useful biomarkers for the management of "post-COVID" patients.

4.1. Strengths and limitations

The strengths of the study reside in the use of next-generation sequencing, without a priori bias, in a relevant number of samples ($n = 50$), which is far larger than similar studies, the use of advanced statistical approaches and the rigorous quality control of clinical data. Nevertheless, our results should be considered in the context of several limitations. First, this was a single-center study. The results should be validated in independent cohorts (beyond the scope of current manuscript). Second, we considered patients with at least moderate to severe diffusion impairment attending to the great prevalence of pulmonary sequelae. Third, potential confounding due to previous comorbidities, clinical management and inpatient complications should not be discarded. Although survivors were matched for age and sex, it is possible that the transcriptomic programs observed reflect unrecognized factors. Fourth, the lack of a non-COVID-19 arm limits the generalization of the findings in the context of ARDS. Fifth, we could not infer whether gene expression disturbances in the blood are due to specific tissue damage or repair mechanisms. The conclusions and relationship derived from our analyses show associations. Additional studies based on in vitro and in vivo models are fundamental to analyze causality. Finally, the identification of the molecular mechanisms associated with post-acute pulmonary sequelae in epithelial, immune and endothelial lung cells would provide valuable information for the development of therapeutic approaches. Although blood cells have been proposed as sentinels reflecting physiological and pathological events occurring in different tissues, whether whole blood can be used as surrogate biopsy material to analyze pathophysiological processes in the local lung environment should be further explored.

5. Conclusions

In summary, post-infection pulmonary dysfunction is associated with a distinct whole-blood transcriptomic pattern in survivors of SARS-CoV-2-driven ARDS after hospital discharge. The transcriptional programs identified provide a valuable opportunity to garner insights into the mechanisms that mediate pulmonary damage and recovery and thus to develop targeted therapeutic interventions and tools for clinical decision-making.

Funding

MCGH is the recipient of a predoctoral fellowship from the University of Lleida. MM is the recipient of a predoctoral fellowship (PFIS: FI21/00187) from Instituto de Salud Carlos III. AC is supported by Instituto de Salud Carlos III (Sara Borrell 2021: CD21/00087). DdGC has received financial support from Instituto de Salud Carlos III (Miguel Servet 2020: CP20/00041), co-funded by the European Social Fund (ESF) "Investing in your future". IML is supported by a Miguel Servet contract (CPII20/00029) from the Instituto de Salud Carlos III, co-funded by the European Social Fund (ESF) "Investing in your future". CIBERES is an initiative of the Instituto de Salud Carlos III. This work is supported by the Instituto de Salud Carlos III (COV20/00110), co-funded by the European Regional Development Fund (ERDF) "A way to make Europe". Supported by: Programa de donaciones "estar preparados"; UNESPA (Madrid, Spain) and Fundación Francisco Soria Melguizo (Madrid, Spain). Funded by: La Fundació La Marató de TV3, project with code 202108–30/−31. COVIDPONENT is funded by the Institut Català de la Salut and Gestió de Serveis Sanitaris. This research was funded in part by a grant (PI19/01805) from the Instituto de Salud Carlos III, co-funded by the European Regional Development Fund (ERDF) "A way to build Europe" and by the Fundación Rioja Salud.

CRediT authorship contribution statement

María C. García Hidalgo: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization. **Rafael Peláez**: Conceptualization, Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Jessica González**: Methodology, Investigation, Resources, Writing – review & editing. **Sally Santisteve:** Investigation. **Iván D. Benítez:** Software, Formal analysis. Marta Molinero: Writing – review & editing. **Manel Pérez-Pons:** Writing – review & editing. **Thalía Belmonte:** Writing – review & editing. **Gerard Torres:** Writing – review & editing. **Anna Moncusí-Moix:** Data curation. **Clara Gort-Paniello:** Data curation. **Maria Aguilà:** Investigation. **Faty Seck:** Investigation. **Paola Carmona:** Investigation. **Jesús Caballero:** Writing – review & editing. **Carme Barberà**: Writing – review & editing. **Jose Ángel Lorente-Balanza:** Writing – review & editing. **Rosario Menéndez:** Writing – review & editing. **Ana Motos:** Writing – review & editing. **Oscar Peñuelas:** Writing – review & editing. **Jordi Riera:** Writing – review & editing. **Jesús F. Bermejo-Martin:** Writing – review & editing. **Antoni Torres:** Writing – review & editing, Funding acquisition. **Ferran Barbé:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **David de Gonzalo-Calvo**: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Ignacio M. Larráyoz:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgments

The authors are indebted to María Arguimbau, Raquel Campo, Natalia Jarillo, Javier Muñoz and Manuel Sánchez (†) for their extensive support with project management and article preparation. This work was supported by IRBLleida Biobank (B.0000682) and "Plataforma Biobancos PT20/00021". The human sample manipulation was performed in the Cell Culture Technical Scientific Service of the Universitat de Lleida (Lleida, Catalonia, Spain).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopha.2022.113617.

References

[1] S.G. Chu, S.P. de Frias, B.A. Raby, I.O. Rosas, An RNA-seq primer for pulmonologists, Eur. Respir. J. 55 (2020) 1–6, https://doi.org/10.1183/13993003.01625-2018.

[2] S. Kilari, A. Sharma, C. Zhao, A. Singh, C. Cai, M. Simeon, A.J. van Wijnen, S. Misra, Identification of novel therapeutic targets for contrast induced acute kidney injury (CI-AKI): alpha blockers as a therapeutic strategy for CI-AKI, Transl. Res. 235 (2021) 32–47, https://doi.org/10.1016/j.trsl.2021.03.005.

[3] R.K. Gupta, J. Rosenheim, L.C. Bell, A. Chandran, J.A. Guerra-Assuncao, G. Pollara, M. Whelan, J. Artico, G. Joy, H. Kurdi, D.M. Altmann, R.J. Boyton, M.K. Maini, A. McKnight, J. Lambourne, T. Cutino-Moguel, C. Manisty, T.A. Treibel, J.C. Moon, B.M. Chain, M. Noursadeghi, Blood transcriptional biomarkers of acute viral infection for detection of pre-symptomatic SARS-CoV-2 infection: a nested, case-control diagnostic accuracy study. Lancet Microbe 2 (2021) e508–e517, https://doi.org/10.1016/S2666-5247(21)00146-4.

[4] S. Mohr, C.C. Liew, The peripheral-blood transcriptome: new insights into disease and risk assessment, Trends Mol. Med. 13 (2007) 422–432, https://doi.org/10.1016/j.molmed.2007.08.003.

[5] J. Atallah, M.K. Mansour, Implications of using host response-based molecular diagnostics on the management of bacterial and viral infections: a review, Front. Med. 9 (2022), 805107, https://doi.org/10.3389/fmed.2022.805107.

[6] M. Hecker, Blood transcriptome profiling captures dysregulated pathways and response to treatment in neuroimmunological disease, EBioMedicine 49 (2019) 2–3, https://doi.org/10.1016/j.ebiom.2019.10.035.

[7] C.J. Rhodes, P. Otero-Núñez, J. Wharton, E.M. Swietlik, S. Kariotis, L. Harbaum, M. J. Dunning, J.M. Elinoff, N. Errington, A.A.R. Thompson, J. Iremonger, J. G. Coghlan, P.A. Corris, L.S. Howard, D.G. Kiely, C. Church, J. Pepke-Zaba, M. Toshner, S.J. Wort, A.A. Desai, M. Humbert, W.C. Nichols, L. Southgate, D. A. Trégouët, R.C. Trembath, I. Prokopenko, S. Gräf, N.W. Morrell, D. Wang, A. Lawrie, M.R. Wilkins, Whole-blood RNA profiles associated with pulmonary arterial hypertension and clinical outcome, Am. J. Respir. Crit. Care Med. 202 (2020) 586–594, https://doi.org/10.1164/rccm.202003-0510OC.

[8] L.A. Huppert, M.A. Matthay, L.B. Ware, Pathogenesis of acute respiratory distress syndrome, Semin. Respir. Crit. Care Med. 40 (2019) 31, https://doi.org/10.1055/S-0039-1683996.

[9] B.T. Thompson, R.C. Chambers, K.D. Liu, Acute respiratory distress syndrome, N. Engl. J. Med. 377 (2017) 562–572, https://doi.org/10.1056/NEJMRA1608077.

[10] M.S. Herridge, M. Moss, C.L. Hough, R.O. Hopkins, T.W. Rice, O.J. Bienvenu, E. Azoulay, Recovery and outcomes after the acute respiratory distress syndrome (ARDS) in patients and their family caregivers, Intensive Care Med 42 (2016) 725–738, https://doi.org/10.1007/s00134-016-4321-8.

[11] T.A. Neff, R. Stocker, H.R. Frey, S. Stein, E.W. Russi, Long-term assessment of lung function in survivors of severe ARDS, Chest 123 (2003) 845–853, https://doi.org/10.1378/chest.123.3.845.

[12] M.F. Mart, L.B. Ware, The long-lasting effects of the acute respiratory distress syndrome, Expert Rev. Respir. Med 14 (2020) 577–586, https://doi.org/10.1080/17476348.2020.1743182.

[13] J. González, I.D. Benítez, P. Carmona, S. Santisteve, A. Monge, A. Moncusí-Moix, C. Gort-Paniello, L. Pinilla, A. Carratalá, M. Zuil, R. Ferrer, A. Ceccato, L. Fernández, A. Motos, J. Riera, R. Menéndez, D. Garcia-Gasulla, O. Peñuelas, J. F. Bermejo-Martin, G. Labarca, J. Caballero, G. Torres, D. de Gonzalo-Calvo, A. Torres, F. Barbé, Pulmonary function and radiologic features in survivors of critical COVID-19, Chest 160 (2021) 187–198, https://doi.org/10.1016/j.chest.2021.02.062.

[14] S. Guler, L. Ebner, C. Aubry-Beigelman, P. Bridevaux, M. Brutsche, C. Clarenbach, C. Garzoni, T. Geiser, A. Lenoir, M. Mancinetti, B. Naccini, S. Ott, L. Piquilloud, M. Prella, Y. Que, P. Soccal, C. von Garnier, M. Funke-Chambour, Pulmonary function and radiological features 4 months after COVID-19: first results from the national prospective observational Swiss COVID-19 lung study, Eur. Respir. J. 57 (2021) 2003690, https://doi.org/10.1183/13993003.03690-2020.

[15] J.R. Blanco, M.J. Cobos-Ceballos, F. Navarro, I. Sanjoaquin, F. Arnaiz de Las Revillas, E. Bernal, L. Buzon-Martin, M. Viribay, L. Romero, S. Espejo-Pérez, B. Valencia, D. Ibañez, D. Ferrer-Pargada, D. Malia, F.G. Gutierrez-Herrero, J. Olalla, B. Jurado-Gamez, J. Ugedo, Pulmonary long-term consequences of COVID-19 infections after hospital discharge, Clin. Microbiol. Infect. 27 (2021) 892–896, https://doi.org/10.1016/j.cmi.2021.02.019.

[16] M.C. García-Hidalgo, J. González, I.D. Benítez, P. Carmona, S. Santisteve, A. Moncusí-Moix, C. Gort-Paniello, F. Rodríguez-Jara, M. Molinero, M. Perez-Pons, G. Torres, J. Caballero, C. Barberà, A.P. Tedim, R. Almansa, A. Ceccato, L. Fernández-Barat, R. Ferrer, D. Garcia-Gasulla, R. Menéndez, A. Motos, O. Peñuelas, J. Riera, J.F. Bermejo-Martin, A. Torres, F. Barbé, D. de Gonzalo-Calvo, Proteomic profiling of lung diffusion impairment in the recovery stage of SARS-CoV-2-induced ARDS, Clin. Transl. Med. 12 (2022), e838, https://doi.org/10.1002/ctm2.838.

[17] A. Torres, M. Arguimbau, J. Bermejo-Martín, R. Campo, A. Ceccato, L. Fernandez-Barat, R. Ferrer, N. Jarillo, J.Á. Lorente-Balanza, R. Menéndez, A. Motos, J. Muñoz, Ó. Peñuelas Rodríguez, R. Pérez, J. Riera, A. Rodríguez, M. Sánchez, F. Barbe, CIBERESUCICOVID: un proyecto estratégico para una mejor comprensión y manejo clínico de la COVID-19 en pacientes críticos, Arch. Bronconeumol. 57 (2021) 1–2, https://doi.org/10.1016/J.ARBRES.2020.10.021.

[18] V.M. Ranieri, G.D. Rubenfeld, B.T. Thompson, N.D. Ferguson, E. Caldwell, E. Fan, L. Camporota, A.S. Slutsky, Acute respiratory distress syndrome: The Berlin definition, JAMA - J. Am. Med. Assoc. 307 (2012) 2526–2533, https://doi.org/10.1001/jama.2012.5669.

[19] R. Pellegrino, G. Viegi, V. Brusasco, R.O. Crapo, F. Burgos, R. Casaburi, A. Coates, C.P.M. van der Grinten, P. Gustafsson, J. Hankinson, R. Jensen, D.C. Johnson, N. MacIntyre, R. McKay, M.R. Miller, D. Navajas, O.F. Pedersen, J. Wanger, Interpretative strategies for lung function tests, Eur. Respir. J. 26 (2005) 948–968, https://doi.org/10.1183/09031936.05.00035205.

[20] E.A. Stuart, Matching methods for causal inference: a review and a look forward, Stat. Sci. 25 (2010) 1–21, https://doi.org/10.1214/09-STS313.

[21] A.I. Oca, Á. Pérez-Sala, A. Pariente, R. Ochoa, S. Velilla, R. Peláez, I.M. Larráyoz, Predictive biomarkers of age-related macular degeneration response to Anti-VEGF treatment. J. Pers. Med. 11 (2021) https://doi.org/10.3390/jpm11121329.

[22] I.M. Larráyoz, A. de Luis, O. Rúa, S. Velilla, J. Cabello, A. Martínez, Molecular effects of doxycycline treatment on pterygium as revealed by massive transcriptome sequencing, PLoS One 7 (2012), e39359, https://doi.org/10.1371/journal.pone.0039359.

[23] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.

[24] H. Varet, L. Brillet-Guéguen, J.-Y. Coppée, M.-A. Dillies, SARTools: A DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data, PLoS One 11 (2016), e0157022, https://doi.org/10.1371/journal.pone.0157022.

[25] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'eustachio, The reactome pathway knowledgebase, Nucleic Acids Res. 48 (2020) D498–D503, https://doi.org/10.1093/nar/gkz1031.

[26] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, M. S. Khodadoust, M.S. Esfahani, B.A. Luca, D. Steiner, M. Diehn, A.A. Alizadeh, Determining cell type abundance and expression from bulk tissues with digital cytometry, Nat. Biotechnol. 37 (2019) 773–782, https://doi.org/10.1038/s41587-019-0114-2.

[27] L. Breiman, Random forests, Mach. Learn 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[28] S.L. Freshour, S. Kiwala, K.C. Cotto, A.C. Coffman, J.F. McMichael, J.J. Song, M. Griffith, O.L. Griffith, A.H. Wagner, Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts, Nucleic Acids Res 49 (2021) D1144–D1151, https://doi.org/10.1093/nar/gkaa1084.

[29] E. Frank, M.A. Hall, I.H. Witten, The WEKA Workbench, in: Data Min. Pract. Mach. Learn. Tools Tech., Fourth edi, 2016.

[30] C.-C. Liew, J. Ma, H.-C. Tang, R. Zheng, A.A. Dempsey, The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J. Lab. Clin. Med. 147 (2006) 126–132, https://doi.org/10.1016/j.lab.2005.10.005.

[31] M.C. García-Hidalgo, J. González, I.D. Benítez, P. Carmona, S. Santisteve, M. Pérez-Pons, A. Moncusí-Moix, C. Gort-Paniello, F. Rodríguez-Jara, M. Molinero, T. Belmonte, G. Torres, G. Labarca, E. Nova-Lamperti, J. Caballero, J.F. Bermejo-Martin, A. Ceccato, L. Fernández-Barat, R. Ferrer, D. Garcia-Gasulla, R. Menéndez, A. Motos, O. Peñuelas, J. Riera, A. Torres, F. Barbé, D. de Gonzalo-Calvo, Identification of circulating microRNA profiles associated with pulmonary function and radiologic features in survivors of SARS-CoV-2-induced ARDS, Emerg. Microbes Infect. (2022) 1–67, https://doi.org/10.1080/22221751.2022.2081615.

[32] G. Häcker, A. Bauer, A. Villunger, Apoptosis in activated T cells: what are the triggers, and what the signal transducers? Cell Cycle 5 (2006) 2421–2424, https://doi.org/10.4161/cc.5.21.3397.

[33] D. Hildeman, T. Jorgensen, J. Kappler, P. Marrack, Apoptosis and the homeostatic control of immune responses, Curr. Opin. Immunol. 19 (2007) 516–521, https://doi.org/10.1016/j.coi.2007.05.005.

[34] P.D. Hughes, G.T. Belz, K.A. Fortner, R.C. Budd, A. Strasser, P. Bouillet, Apoptosis regulators Fas and Bim cooperate in shutdown of chronic immune responses and prevention of autoimmunity, Immunity 28 (2008) 197–205, https://doi.org/10.1016/j.immuni.2007.12.017.

[35] N. Andina, S. Conus, E.M. Schneider, M.F. Fey, H.U. Simon, Induction of Bim limits cytokine-mediated prolonged survival of neutrophils, Cell Death Differ. 16 (2009) 1248–1255, https://doi.org/10.1038/cdd.2009.50.

[36] S.S. Burgener, N.G.F. Leborgne, S.J. Snipas, G.S. Salvesen, P.I. Bird, C. Benarafa, Cathepsin G inhibition by Serpinb1 and Serpinb6 prevents programmed necrosis in neutrophils and monocytes and reduces GSDMD-driven inflammation, Cell Rep. 27 (2019) 3646–3656.e5, https://doi.org/10.1016/j.celrep.2019.05.065.

[37] S. Liang, F. Wang, C. Bao, J. Han, Y. Guo, F. Liu, Y. Zhang, BAG2 ameliorates endoplasmic reticulum stress-induced cell apoptosis in Mycobacterium tuberculosis-infected macrophages through selective autophagy, Autophagy 16 (2020) 1453–1467, https://doi.org/10.1080/15548627.2019.1687214.

[38] S. Liang, Z. Song, Y. Wu, Y. Gao, M. Gao, F. Liu, F. Wang, Y. Zhang, MicroRNA-27b modulates inflammatory response and apoptosis during mycobacterium tuberculosis infection, J. Immunol. 200 (2018) 3506–3518, https://doi.org/10.4049/jimmunol.1701448.

[39] N. Zhang, C. Zhao, X. Zhang, X. Cui, Y. Zhao, J. Yang, X. Gao, Growth arrest-specific 2 protein family: structure and function, Cell Prolif. 54 (2021), e12934, https://doi.org/10.1111/cpr.12934.

[40] H.-C. Hsu, D.K. Scott, J.D. Mountz, Impaired apoptosis and immune senescence - cause or effect? Immunol. Rev. 205 (2005) 130–146, https://doi.org/10.1111/j.0105-2896.2005.00270.x.

[41] C.M. Weyand, J.J. Goronzy, Aging of the immune system. mechanisms and therapeutic targets, Ann. Am. Thorac. Soc. 13 Suppl 5 (2016) S422–S428, https://doi.org/10.1513/AnnalsATS.201602-095AW.

[42] M.P. Baar, R.M.C. Brandt, D.A. Putavet, J.D.D. Klein, K.W.J. Derks, B.R. M. Bourgeois, S. Stryeck, Y. Rijksen, H. van Willigenburg, D.A. Feijtel, I. van der

Pluijm, J. Essers, W.A. van Cappellen, W.F. van IJcken, A.B. Houtsmuller, J. Pothof, R.W.F. de Bruin, T. Madl, J.H.J. Hoeijmakers, J. Campisi, P.L.J. de Keizer, Targeted apoptosis of senescent cells restores tissue homeostasis in response to chemotoxicity and aging, Cell 169 (2017) 132–147.e16, https://doi.org/10.1016/j.cell.2017.02.031.

[43] K. Sobierajska, W.M. Ciszewski, M.E. Wawro, K. Wieczorek-Szukała, J. Boncela, I. Papiewska-Pajak, J. Niewiarowska, M.A. Kowalska, TUBB4B downregulation is critical for increasing migration of metastatic colon cancer cells, Cells 8 (2019), https://doi.org/10.3390/cells8080810.

[44] S.H. Cant, J.A. Pitcher, G protein-coupled receptor kinase 2-mediated phosphorylation of ezrin is required for G protein-coupled receptor-dependent reorganization of the actin cytoskeleton, Mol. Biol. Cell. 16 (2005) 3088–3099, https://doi.org/10.1091/mbc.e04-10-0877.

[45] A. Sgorbissa, R. Benetti, S. Marzinotto, C. Schneider, C. Brancolini, Caspase-3 and caspase-7 but not caspase-6 cleave Gas2 in vitro: implications for microfilament reorganization during apoptosis. J. Cell Sci. 112 (Pt 2) (1999) 4475–4482, https://doi.org/10.1242/jcs.112.23.4475.

[46] M.J. Stroud, A. Nazgiewicz, E.A. McKenzie, Y. Wang, R.A. Kammerer, C. Ballestrem, GAS2-like proteins mediate communication between microtubules and actin through interactions with end-binding proteins, J. Cell Sci. 127 (2014) 2672–2682, https://doi.org/10.1242/jcs.140558.

[47] T.I. Arnon, Y. Xu, C. Lo, T. Pham, J. An, S. Coughlin, G.W. Dorn, J.G. Cyster, GRK2-dependent S1PR1 desensitization is required for lymphocytes to overcome their attraction to blood, Science 333 (2011) 1898–1903, https://doi.org/10.1126/science.1208248.

[48] F.R. Villagomez, J.D. Diaz-Valencia, E. Ovalle-García, A. Antillón, I. Ortega-Blake, H. Romero-Ramírez, J.F. Cerna-Cortes, R. Rosales-Reyes, L. Santos-Argumedo, G. Patiño-López, TBC1D10C is a cytoskeletal functional linker that modulates cell spreading and phagocytosis in macrophages, Sci. Rep. 11 (2021) 20946, https://doi.org/10.1038/s41598-021-00450-z.

[49] L.M. Crosby, C.M. Waters, Epithelial repair mechanisms in the lung, Am. J. Physiol. Lung Cell. Mol. Physiol. 298 (2010) L715–L731, https://doi.org/10.1152/ajplung.00361.2009.

[50] R.L. Zemans, N. Briones, M. Campbell, J. McClendon, S.K. Young, T. Suzuki, I. V. Yang, S. De Langhe, S.D. Reynolds, R.J. Mason, M. Kahn, P.M. Henson, S. P. Colgan, G.P. Downey, Neutrophil transmigration triggers repair of the lung epithelium via β-catenin signaling, Proc. Natl. Acad. Sci. 108 (2011) 15990–15995, https://doi.org/10.1073/pnas.1110144108.

[51] S. Kavela, S.R. Shinde, R. Ratheesh, K. Viswakalyan, M.D. Bashyam, S. Gowrishankar, M. Vamsy, S. Pattnaik, S. Rao, R.A. Sastry, M. Srinivasulu, J. Chen, S. Maddika, PNUTS functions as a proto-oncogene by sequestering PTEN, Cancer Res 73 (2013) 205–214, https://doi.org/10.1158/0008-5472.CAN-12-1394.

[52] Y. Lan, J. Lou, J. Hu, Z. Yu, W. Lyu, B. Zhang, Downregulation of SNRPG induces cell cycle arrest and sensitizes human glioblastoma cells to temozolomide by targeting Myc through a p53-dependent signaling pathway, Cancer Biol. Med. 17 (2020) 112–131, https://doi.org/10.20892/j.issn.2095-3941.2019.0164.

[53] L. Qin, J. Guo, Q. Zheng, H. Zhang, BAG2 structure, function and involvement in disease, Cell. Mol. Biol. Lett. 21 (2016) 18, https://doi.org/10.1186/s11658-016-0020-2.

[54] H.M. Froggatt, A.T. Harding, R.R. Chaparian, N.S. Heaton, ETV7 limits antiviral gene expression and control of influenza viruses, Sci. Signal. 14 (2021), https://doi.org/10.1126/scisignal.abe1194.

[55] E.L. Robinson, A.H. Baker, M. Brittan, I. McCracken, G. Condorelli, C. Emanueli, K. P. Srivastava, C. Gaetano, T. Thum, M. Vanhaverbeke, C. Angione, S. Heymans, Y. Devaux, T. Pedrazzini, F. Martelli, Dissecting the transcriptome in cardiovascular disease, Cardiovasc. Res. 118 (2021) 1004–1019, https://doi.org/10.1093/cvr/cvab117.