iScience, Volume *23*

Supplemental Information

SARS-CoV-2 Infection Boosts *MX1* Antiviral Effector in COVID-19 Patients

Juan Bizzotto, Pablo Sanchis, Mercedes Abbate, Sofía Lage-Vickers, Rosario Lavignolle, Ayelén Toro, Santiago Olszevicki, Agustina Sabater, Florencia Cascardo, Elba Vazquez, Javier Cotignola, and Geraldine Gueron

Table S1. Patient demographics for the GSE152075 dataset. Related to all figures.

		non-COVID-19	COVID-19	*P*-value
Number of patients		50	403	
Sex	**Male**	22 (44%)	166 (41.2%)	
	Female	28 (56%)	187 (46.4%)	0.763 [a]
	Unknown	-	50 (12.4%)	
Age (years)	**Range**	12 - 91	2 - 91	
	Inter Quartile Range	29-63	41-71	
	Median	46.5	56	
	Media	46.5	55.6	0.002 [b]
	<30s	13 (26%)	41 (10.2%)	
	30s	5 (10%)	51 (12.7%)	
	40s	10 (20%)	55 (13.6%)	
	50s	7 (14%)	80 (19.9%)	
	60s	8 (16%)	50 (12.4%)	
	≥70	7 (14%)	110 (27.3%)	0.012 [c]
	Unknown	-	16 (3.9%)	
Viral load	**High (Ct <19)**	-	106 (26.3%)	
	Mid (Ct 19-24)	-	197 (48.9%)	
	Low (Ct >24)	-	84 (20.9%)	
	Unknown	-	16 (3.9%)	

Table contains the number and % of COVID-19 and non-COVID-19 patients according to sex, age and viral load.[a] Fisher's exact; [b] t-Student test; [c] Chi-square test.

Table S2. Host receptor and antiviral gene correlation in all patients (Global), non-COVID-19 patients, or COVID-19 patients. Related to Figure 3A.

Genes		Global		Negative (non-COVID-19)		Positive (COVID-19)	
		r	*P*-value	r	*P*-value	r	*P*-value
ACE2	ADAM17	-0.013	0.786	0.138	0.340	-0.008	0.878
	BSG	0.045	0.343	0.093	0.522	**0.139**	**0.005**
	CTSB	0.177	$1.59x10^{-04}$	0.170	0.237	**0.267**	**$5.22x10^{-08}$**
	CTSL	0.291	$2.85x10^{-10}$	0.367	0.009	0.284	$6.19x10^{-09}$
	HIF1A	0.233	$5.61x10^{-07}$	0.146	0.312	**0.233**	**$2.28x10^{-06}$**
	HMOX1	0.180	$1.15x10^{-04}$	0.053	0.714	**0.210**	**$2.22x10^{-05}$**
	IRF3	0.050	0.284	0.178	0.217	0.086	0.084
	MX1	0.439	$<1x10^{-15}$	0.132	0.361	**0.416**	**$<1x10^{-15}$**
	MX2	0.402	$<1x10^{-15}$	0.168	0.244	**0.359**	**$1.07x10^{-13}$**
	NRF2	0.291	$2.82x10^{-10}$	0.090	0.534	**0.360**	**$8.22x10^{-14}$**
ADAM17	NRF2	0.189	$5.17x10^{-05}$	0.207	0.150	**0.187**	**$1.57x10^{-04}$**
BSG	ADAM17	0.182	$9.70x10^{-05}$	0.106	0.464	**0.195**	**$8.27x10^{-05}$**
	HIF1A	0.313	$1.01x10^{-11}$	0.176	0.221	**0.366**	**$3.26x10^{-14}$**
	HMOX1	0.510	$<1x10^{-15}$	0.398	0.004	0.531	$<1x10^{-15}$
	IRF3	0.390	$<1x10^{-15}$	0.504	$1.92x10^{-04}$	0.350	$5.01x10^{-13}$
	MX1	0.270	$5.52x10^{-09}$	0.461	0.001	0.412	$<1x10^{-15}$
	MX2	0.200	$1.76x10^{-05}$	0.072	0.617	**0.379**	**$3.55x10^{-15}$**
	NRF2	0.483	$<1x10^{-15}$	0.590	$6.59x10^{-06}$	0.466	$<1x10^{-15}$
CTSB	ADAM17	0.212	$5.07x10^{-06}$	0.008	0.954	**0.232**	**$2.55x10^{-06}$**
	BSG	0.674	$<1x10^{-15}$	0.707	$9.88x10^{-09}$	0.623	$<1x10^{-15}$
	CTSL	0.474	$<1x10^{-15}$	0.686	$3.91x10^{-08}$	0.492	$<1x10^{-15}$
	HIF1A	0.509	$<1x10^{-15}$	0.281	0.048	0.572	$<1x10^{-15}$
	HMOX1	0.579	$<1x10^{-15}$	0.466	0.001	0.590	$<1x10^{-15}$
	IRF3	0.475	$<1x10^{-15}$	0.613	$2.26x10^{-06}$	0.437	$<1x10^{-15}$
	MX1	0.499	$<1x10^{-15}$	0.465	0.001	0.657	$<1x10^{-15}$
	MX2	0.400	$<1x10^{-15}$	-0.012	0.933	**0.595**	**$<1x10^{-15}$**
	NRF2	0.685	$<1x10^{-15}$	0.575	$1.24x10^{-05}$	0.698	$<1x10^{-15}$

Genes		Global		Negative (non-COVID-19)		Positive (COVID-19)	
		r	***P*-value**	**r**	***P*-value**	**r**	***P*-value**
CTSL	ADAM17	0.169	$2.91x10^{-04}$	0.303	0.033	0.162	0.001
	BSG	0.237	$3.31x10^{-07}$	0.599	$4.41x10^{-06}$	0.237	$1.56x10^{-06}$
	HIF1A	0.526	$<1x10^{-15}$	0.261	0.067	**0.551**	**$<1x10^{-15}$**
	HMOX1	0.469	$<1x10^{-15}$	0.412	0.003	0.477	$<1x10^{-15}$
	IRF3	0.297	$1.14x10^{-10}$	0.691	$2.85x10^{-08}$	0.267	$5.37x10^{-08}$
	MX1	0.452	$<1x10^{-15}$	0.507	$1.71x10^{-04}$	0.456	$<1x10^{-15}$
	MX2	0.427	$<1x10^{-15}$	0.262	0.066	**0.450**	**$<1x10^{-15}$**
	NRF2	0.473	$<1x10^{-15}$	0.555	$2.94x10^{-05}$	0.480	$<1x10^{-15}$
HIF1A	ADAM17	0.189	$5.40x10^{-05}$	0.112	0.438	**0.203**	**$4.09x10^{-05}$**
	NRF2	0.687	$<1x10^{-15}$	0.544	$4.39x10^{-05}$	0.713	$<1x10^{-15}$
HMOX1	ADAM17	0.157	0.001	0.039	0.786	**0.165**	**0.001**
	HIF1A	0.578	$<1x10^{-15}$	0.402	0.004	0.598	$<1x10^{-15}$
	IRF3	0.301	$5.88x10^{-11}$	0.358	0.011	0.288	$4.03x10^{-09}$
	NRF2	0.550	$<1x10^{-15}$	0.322	0.023	0.565	$<1x10^{-15}$
IRF3	ADAM17	0.257	$2.99x10^{-08}$	0.303	0.032	0.247	$5.21x10^{-07}$
	HIF1A	0.249	$7.95x10^{-08}$	0.175	0.223	**0.271**	**$3.11x10^{-08}$**
	NRF2	0.388	$<1x10^{-15}$	0.585	$8.01x10^{-06}$	0.353	$2.8x10^{-13}$
MX1	ADAM17	0.146	0.002	0.312	0.028	0.177	$3.64x10^{-04}$
	HIF1A	0.593	$<1x10^{-15}$	0.465	0.001	0.623	$<1x10^{-15}$
	HMOX1	0.432	$<1x10^{-15}$	0.429	0.002	0.470	$<1x10^{-15}$
	IRF3	0.250	$7.28x10^{-08}$	0.658	$2.08x10^{-07}$	0.288	$3.90x10^{-09}$
	MX2	0.897	$<1x10^{-15}$	0.642	$4.97x10^{-07}$	0.899	$<1x10^{-15}$
	NRF2	0.652	$<1x10^{-15}$	0.593	$5.66x10^{-06}$	0.745	$<1x10^{-15}$
MX2	ADAM17	0.189	$4.94x10^{-05}$	0.363	0.01	0.226	$4.48x10^{-06}$
	HIF1A	0.624	$<1x10^{-15}$	0.417	0.003	0.666	$<1x10^{-15}$
	HMOX1	0.362	$1.78x10^{-15}$	0.079	0.587	**0.423**	**$<1x10^{-15}$**
	IRF3	0.235	$4.42x10^{-07}$	0.277	0.051	**0.319**	**$5.61x10^{-11}$**
	NRF2	0.580	$<1x10^{-15}$	0.486	$<1x10^{-15}$	0.688	$<1x10^{-15}$

Genes		Global		Negative (non-COVID-19)		Positive (COVID-19)	
		r	*P*-value	r	*P*-value	r	*P*-value
TMPRSS2	ACE2	0.138	0.003	0.182	0.206	**0.220**	**$8.66x10^{-06}$**
	ADAM17	0.291	$2.68x10^{-10}$	0.168	0.243	**0.296**	**$1.43x10^{-09}$**
	BSG	0.432	$<1x10^{-15}$	0.534	$6.55x10^{-05}$	0.364	$4.71x10^{-14}$
	CTSB	0.516	$<1x10^{-15}$	0.475	$4.87x10^{-04}$	0.473	$<1x10^{-15}$
	CTSL	0.264	$1.22x10^{-08}$	0.495	$2.61x10^{-04}$	0.273	$2.67x10^{-08}$
	HIF1A	0.291	$2.64x10^{-10}$	0.221	0.123	**0.324**	**$2.51x10^{-11}$**
	HMOX1	0.313	$8.95x10^{-12}$	0.315	0.026	0.298	$1.03x10^{-09}$
	IRF3	0.431	$<1x10^{-15}$	0.565	$1.95x10^{-05}$	0.390	$4.44x10^{-16}$
	MX1	0.290	$3.35x10^{-10}$	0.679	$5.88x10^{-08}$	0.393	$2.22x10^{-16}$
	MX2	0.266	$9.16x10^{-09}$	0.380	0.006	0.404	$<1x10^{-15}$
	NRF2	0.485	$<1x10^{-15}$	0.549	$3.63x10^{-05}$	0.473	$<1x10^{-15}$

Table contains the r value and *P*-value for gene correlations. Genes that have significant correlation in COVID-19 patients but non-significant correlation in non-COVID-19 are highlighted in bold. Data was taken from the GSE152075 dataset.

TRANSPARENT METHODS

Transcriptome datasets selection and study population

We browsed the Gene Expression Omnibus (GEO) repository (Barrett et al., 2013) using the following keywords and expressions: [(COVID) OR (COVID-19) OR (CORONAVIRUS) OR (SARS-CoV-2) OR (2019-nCoV)] AND [(transcriptomics) OR (RNA-seq) OR (microarray) OR (expression) OR (transcriptome)]. All potentially relevant datasets were further evaluated in detail by 2-3 authors. The eligibility criteria included: (i) publicly available transcriptome data; (ii) detailed sample information; (iii) detailed protocol information; (iv) ≥ 60 samples.

We selected the GSE152075 dataset, which contains RNA-seq data from 430 SARS-CoV-2 positive and 54 negative patients (Lieberman et al., 2020). RNA was isolated from nasopharyngeal swabs and sequenced in an Illumina NextSeq 500 instrument. Reads were

pseudoaligned to the human transcriptome and pre-processed by the authors of the original study (Lieberman et al., 2020). Clinico-pathological information included age, sex, and viral load (expressed as cycle threshold (Ct) by RT-qPCR for the N1 viral gene at time of diagnosis). The interpretation for viral load was: the higher the viral load, the lower de Ct.

RNA-seq analysis

We downloaded the pseudoaligned and pre-processed RNA-seq data. We removed samples with >70% of total genes with 0 sequences reads considering them as very low-quality samples that might introduce a bias. Normalization, batch effect correction and differential expression were performed with R package DEseq2 v1.28.1 (Love et al., 2014).

Statistical analyses

Wilcoxon rank sum test was performed to determine statistical differences between categorical groups. Age was categorized according to the WHO guidelines ("World Health Organization 2020 - Novel Coronavirus (2019-nCoV) Situation Report - 1," n.d.): <30 years old, every 10 years between 30-70 years old and ≥70 years old. Two-sided, increasing and decreasing Jonckheere-Terpstra trend tests (with 500 permutations) were used to determine statistical trends between gene expression and age groups. To study pairwise correlations between continuous variables, Spearman's rank correlation coefficient was calculated. Multilinear regression analyses were performed to determine the correlation between the expression of two genes and viral load. To estimate the regression coefficients of the different models, we used a multivariable regression including gene expression, viral load and age as covariates.

All statistical analyses were done in Stata v14 (StataCorp LLC, College Station, TX, USA) or GraphPad Prism (La Jolla, CA, USA). Statistical significance was set as $P<0.05$. We did not correct *P* value for multiple testing.

Principal Component Analysis (PCA) was done with the factorextra and dplyr packages in R (Kassambara and Mundt, 2020). We included only the selected host receptors and antiviral genes.

All results were plotted using ggplot2 (Wickham and Hadley, 2016), ggpubr (Kassambara, 2020), GGally (Schloerke et al., 2020), canvasXpress (Neuhaus and Brett, 2020) in R, or GraphPad Prism (La Jolla, CA, USA).

SUPLEMENTARY REFERENCES

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 41, D991-5. https://doi.org/10.1093/nar/gks1193

Kassambara, A., 2020. ggpubr: “ggplot2” Based Publication Ready Plots.

Kassambara, A., Mundt, F., 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package.

Lieberman, N.A.P., Peddu, V., Xie, H., Shrestha, L., Huang, M.-L., Mears, M.C., Cajimat, M.N., Bente, D.A., Shi, P.-Y., Bovier, F., Roychoudhury, P., Jerome, K.R., Moscona, A., Porotto, M., Greninger, A.L., 2020. In vivo antiviral host response to SARS-CoV-2 by viral load, sex, and age. bioRxiv Prepr. Serv. Biol. 2020.06.22.165225. https://doi.org/10.1101/2020.06.22.165225

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

Neuhaus, I., Brett, C., 2020. canvasXpress: Visualization Package for CanvasXpress in R.

Schloerke, B., Cook, D., Larmarange, J., Francois, B., Marbach, M., Thoen, E., Elberg, A., Crowley, J., 2020. GGally: Extension to “ggplot2”. R package.

Wickham, Hadley, 2016. ggplot2. Elegant Graphics for Data Analysis. Springer-Verlag, NY.

World Health Organization 2020 - Novel Coronavirus (2019-nCoV) Situation Report - 1 [WWW Document], n.d. URL https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4