



Lung expression of genes putatively involved in SARS-CoV-2 infection is modulated in *cis* by germline variants

Chiara E. Cotroneo¹ · Nunzia Mangano¹ · Tommaso A. Dragani¹ · Francesca Colombo²

Received: 1 July 2020 / Revised: 9 December 2020 / Accepted: 9 February 2021 / Published online: 1 March 2021
© The Author(s), under exclusive licence to European Society of Human Genetics 2021

Abstract

Germline variants in genes involved in SARS-CoV-2 cell entry and in host innate immune responses to viruses may influence the susceptibility to infection. This study used whole-genome analyses of lung tissue to identify polymorphisms acting as expression quantitative trait loci (eQTLs) for 60 genes of relevance to SARS-CoV-2 infection susceptibility. The expression of genes with confirmed or possible roles in viral entry–replication and in host antiviral responses was studied in the non-diseased lung tissue of 408 lung adenocarcinoma patients. No gene was differently expressed by sex, but *APOBEC3H* levels were higher and *PARP12* levels lower in older individuals. A total of 125 *cis*-eQTLs (false discovery rate < 0.05) was found to modulate mRNA expression of 15 genes (*ABO*, *ANPEP*, *AP2A2*, *APOBEC3D*, *APOBEC3G*, *BSG*, *CLEC4G*, *DDX58*, *DPP4*, *FURIN*, *FYCO1*, *RAB14*, *SERINC3*, *TRIM5*, *ZCRB1*). eQTLs regulating *ABO* and *FYCO1* were found in COVID-19 susceptibility loci. No *trans*-eQTLs were identified. Genetic control of the expression of these 15 genes, which encode putative virus receptors, proteins required for vesicle trafficking, enzymes that interfere with viral replication, and other restriction factors, may underlie interindividual differences in risk or severity of infection with SARS-CoV-2 or other viruses.

Introduction

SARS-CoV-2 infection has a wide range of clinical presentations, from asymptomatic to severe or critical, leading in some cases to death [1, 2]. Data from Italy exemplify the great variability in prognosis [3]. For example, in a sample of 705,465 persons who tested positive for SARS-CoV-2 genome (data from December 2020), 58.5% were asymptomatic, 36.5% had few or mild symptoms, and about 5% had severe symptoms requiring hospitalization. Considering all 1,624,269 PCR-confirmed cases in Italy, the case-fatality rate is 3.4%. Prognosis depends on age, with 85.7% of

deaths registered in persons ≥ 70 years old. Prognosis also depends on sex, with a higher case-fatality rate for men than women (4.1% vs. 2.8%, respectively). According to a meta-analysis of more than 3000 patients, male sex and age > 65 years are poor prognostic factors [4]. Why SARS-CoV-2 infection has such variable presentations and why COVID-19 is more fatal in men than women are currently largely unknown. Preexisting medical conditions such as diabetes and hypertension and a smoking habit have been reported to increase the risk of severe disease [4, 5], and there is emerging evidence for the existence of at least two strains of the virus that vary in virulence [6, 7]. Individual genetic constitution may also play a role in determining the severity of a SARS-CoV-2 infection.

The influence of genetics on pulmonary infections has already been observed. For example, susceptibility to and severity of influenza vary according to the genotype of single nucleotide polymorphisms (SNPs) in several genes involved in interactions with the influenza A virus, in inflammatory processes, and in other host responses (reviewed in [8]). These genes include *IFITM3*, a member of the interferon-induced transmembrane protein (IFITM) family, and *TMPRSS2*, which encodes a transmembrane protease whose activity is essential for viral entry. Both

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-021-00831-y>.

✉ Tommaso A. Dragani
tommaso.dragani@istitutotumori.mi.it

¹ Department of Research, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

² Institute of Biomedical Technologies, National Research Council (ITB-CNR), Segrate, MI, Italy

coding and noncoding polymorphisms in these genes have been associated with the risk or severity of influenza A infection [9–11]. These results suggest that not only genetic variants affecting protein function but also those affecting gene expression modulate the response to viral infection. For severe acute respiratory syndrome (SARS), caused by the SARS-CoV virus, the risk or severity of infection has been associated with polymorphisms in genes involved in inflammatory and immune responses [12–14] or viral entry [15]. Recently, a genome-wide association study identified two susceptibility loci (at 3p21.31 and 9q34.2) for developing severe COVID-19 [16].

For the new coronavirus SARS-CoV-2, at least two human proteins so far have been identified as being involved in viral entry and replication in cells. Angiotensin I converting enzyme 2 (ACE2), the main receptor for viral attachment [17, 18], is a transmembrane protein encoded by the *ACE2* gene. According to molecular modeling studies, two coding variants may alter the binding affinity between ACE2 and the viral spike protein (protein S) [19]. SARS-CoV-2 entry also depends on the protease encoded by *TMPRSS2*, which cleaves the viral spike protein, an essential step for adsorption of the virus to the cell [17]. A new SARS-CoV-2 invasion mechanism, involving interactions between the spike protein and CD147, encoded by the *BSG* gene, has recently been proposed [20]. Several other proteins involved in viral genome replication or vesicle formation, as well as receptors for other coronaviruses, have been proposed as factors putatively involved in SARS-CoV-2 infection [21].

In addition to the human proteins needed for viral entry, other proteins involved in viral infection are the so-called antiviral restriction factors. These components of the innate immune response system target viruses and interfere with their life cycle, for instance by modifying their proteins or RNA (reviewed in [22]). Antiviral restriction factors include members of the IFITM family, the APOBEC3 family of RNA editing enzymes, the adenosine deaminase ADAR, and members of the poly(ADP-ribose) polymerase (PARP) family [23–25]. Coding variants in these genes may also explain the different responses to SARS-CoV-2 infection in different people. As already observed for influenza, risk or severity of infection may also depend on noncoding variants that influence gene expression, for instance by modifying the binding sites for transcription factors or altering enhancer sequences [9, 10, 26]. These regulatory genetic variants are termed expression quantitative trait loci (eQTLs) because they alter gene expression in a quantitative manner [27].

Under the hypothesis that individual genetic constitution plays a role in SARS-CoV-2 infection, we examined the expression in lung of 60 genes of relevance to SARS-CoV-2 infection susceptibility. This gene set included genes

encoding proteins with confirmed or possible roles in SARS-CoV2 entry or replication and proteins involved in host antiviral responses, as well as genes in the two COVID-19 susceptibility loci. Taking advantage of an existing data set of genotype and transcriptome data from noninvolved lung tissue from patients with lung adenocarcinoma, we investigated whether the expression of these candidate genes reflected clinical characteristics known to be prognostic of SARS-CoV-2 outcome (i.e., age and sex) or was genetically controlled by *cis* or *trans* regulatory polymorphisms.

Materials and methods

Study design

We selected 60 genes of relevance to SARS-CoV-2 infection (Supplementary Table 1). This gene set included: 24 genes encoding proteins with confirmed or proposed functions in SARS-CoV-2 and other coronavirus adsorption or entry into human cells, viral genome regulation, and vesicle trafficking; 29 genes encoding antiviral restriction factors and immune response-related proteins; and seven genes recently found in severe COVID-19 susceptibility loci. The expression of these genes and the genotypes of germline variants were studied in apparently normal lung tissue from 408 Italian patients who had undergone lobectomy for lung adenocarcinoma. Samples of noninvolved lung tissue had been collected from portions of lung lobes as far as possible from the tumor.

Transcriptome profiling

Lung transcriptome data, obtained using Illumina HumanHT-12 v4 Expression BeadChips, were already available in our lab (GEO database accession numbers: GSE71181 and GSE123352). Microarray raw data were subjected to preprocessing and quality checking using R version 3.6.3. Raw data were log₂-transformed and normalized using the robust spline normalization method implemented in the lumi Bioconductor package [28]. Corrected data were then collapsed from probe level to gene level by selecting, for each gene, the probe with the highest mean intensity across samples.

After preprocessing, measured intensities for genes of interest were extracted and then merged after adjusting for batch effects using the ComBat function from the sva R Bioconductor package [29], using sex and age at surgery of each patient as covariates for the normalization step. The cellular composition of the lung tissue samples was estimated from transcriptome profiles using the bioinformatics tool xCell (<https://xcell.ucsf.edu/>) [30].

Correlation of gene expression with sex and age

Sex and age of the patients were tested as possible linear predictors of the observed \log_2 -transformed transcript level of each gene using a linear regression model, by the `glm` function in R version 3.6.3. Significant correlations were considered when $P < 0.001$.

SNP genotyping

Genotyping data of a subset of 201 patients were already available (described in [31]). Genome-wide genotyping of 200 ng genomic DNA from the remaining 207 patients was carried out using Infinium HumanOmni2.5–8 v1.2 BeadChip microarrays (Illumina). After completion of the assay, the BeadChips were scanned with the two-color confocal Illumina HiScan System at a 0.375 μm pixel resolution. Image intensities were extracted and genotypes were determined using Illumina's BeadStudio software.

Genotype data were subjected to quality control using PLINK v1.90b6.16 [32] following the protocol described in [33]. All individuals whose heterozygosity rates for markers on chromosome X were inconsistent with their biological sex were discarded. Other discarded samples corresponded to individuals where 3% of the total genotyped SNPs were missing and whose heterozygosity rates deviated by three or more standard deviations from the average heterozygosity rate of the whole study group. In the remaining samples, all genotyped SNPs with a call rate $< 95\%$, a minor allele frequency (MAF) $< 2\%$, or a Hardy–Weinberg equilibrium $P < 1.0 \times 10^{-7}$ were excluded from further analyses. Probes mapping to multiple regions were also removed, as were mitochondrial SNPs and duplicate SNPs (i.e., those with duplicate IDs or genomic positions). The resulting data set was then converted to a tabular format with the option `--recode A-transpose`.

eQTL analyses

To test whether gene expression in lung tissue varied with the genotype of SNPs, we carried out an eQTL analysis by standard additive linear regression model, using the MatrixEQTL package [34] in R environment, assuming that genotype has an additive effect on gene expression. Briefly, genotypes were expressed as integers (0, 1, or 2) according to the number of minor alleles at each SNP. Sex and age at surgery of each patient were used as covariates. eQTLs were considered *in cis* where the corresponding SNP was located within one megabase of the gene mapping coordinates, and *in trans* for all other cases. Correction of P -values for multiple testing was carried out using the Benjamini–Hochberg method [35] to obtain the false discovery rate (FDR). Genome-wide significance threshold for

the analysis was set at $\text{FDR} < 0.05$. Differences in expression levels (\log_2 -transformed values) among the three genotype groups were tested for significance by one-way ANOVA followed by Tukey's test for multiple comparisons ($P < 0.05$ was considered as statistically significant).

To understand the tissue specificity of eQTLs, we searched the Genotype-Tissue Expression (GTEx) database [<https://www.gtexportal.org/home/>, GTEx Analysis Release V8, dbGaP Accession phs000424.v8.p2, accessed on December, 2020]. All identified *cis*-eQTLs were singularly searched using the variant browser, and we looked for our target genes among the single-tissue eQTL results. In particular, data (normalized effect size and P value) of lung eQTLs were reported.

Linkage disequilibrium data and minor allele frequencies in the different populations were retrieved from the Ensembl genome browser. To investigate potential functional roles of the top significant SNP of each eQTL, we used the SNP Function Prediction tool at the SNPinfo Web Server [36], with the default settings.

Results

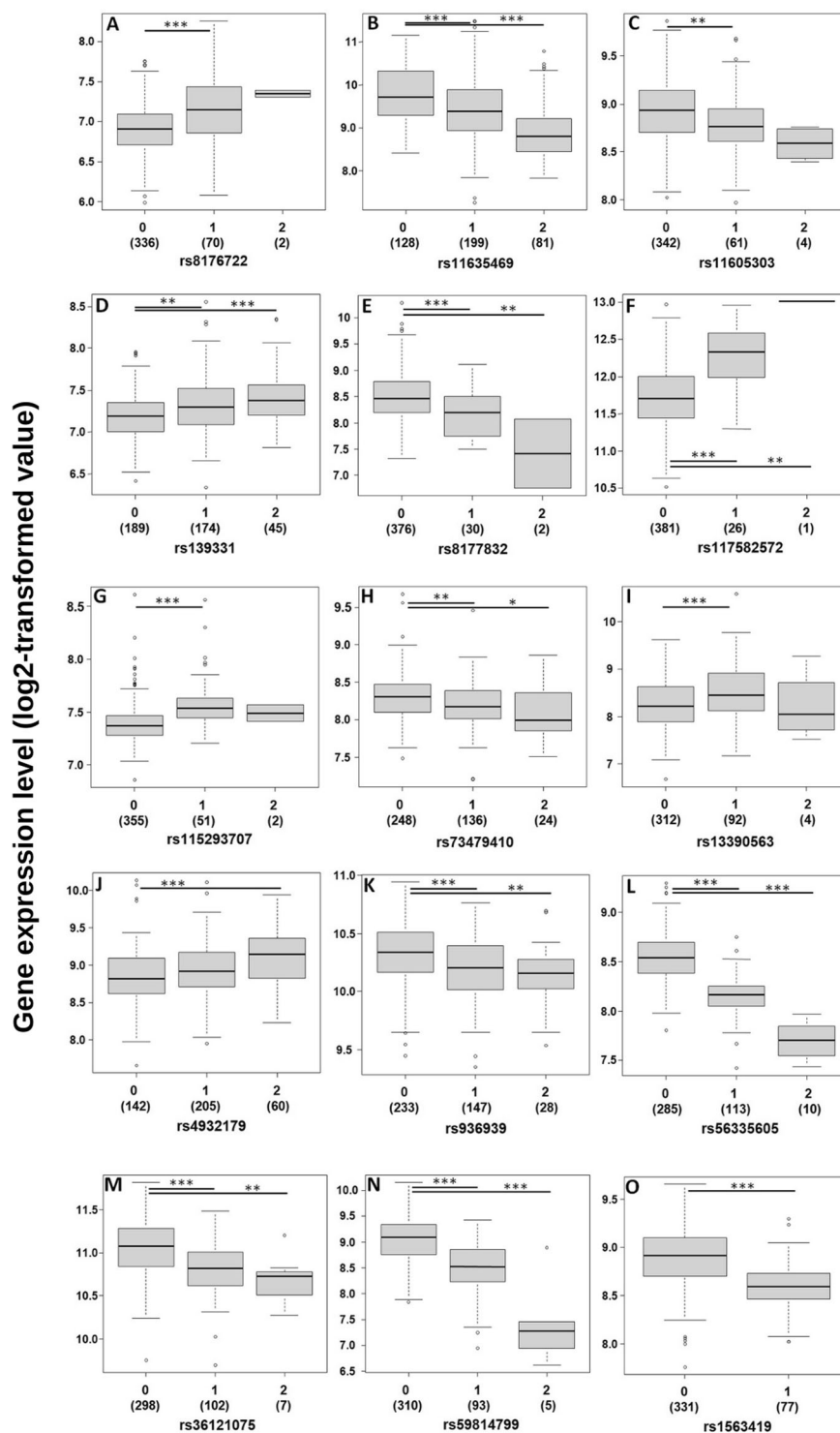
This study considered transcriptomic and genotype data from samples of apparently normal lung tissue from 408 patients with lung adenocarcinoma (Supplementary Table 2). The study group had a median age of 65 years (range, 36–84) and a predominance of men (67.2%). The vast majority (88.2%) were ever smokers (i.e., current and former smokers) and 64.6% had pathological stage I disease.

Transcription profiles of lung tissue samples were used to estimate the cellular composition. Among 64 possible cell types, 7 myeloid, 2 lymphoid, 2 epithelial, 6 stromal, and 3 hematopoietic stem cell types were found to be enriched in at least 80% of patients (not shown). These findings indicate that the expression data used in this study belong to the different cell types normally present in lung tissue, and are not just from an immune infiltrate.

Effect of sex and age on gene expression

Of the 60 selected genes, all were expressed at detectable levels in lung tissue with the exception of *IFITM1*, which did not pass preprocessing and quality checking filters. A first analysis assessed whether the expression levels of the remaining 59 genes expressed in lung associated with two clinical variables known to affect prognosis in COVID-19, i.e., sex and age (Supplementary Table 1). No gene was differently expressed between men and women. *PARP12* mRNA levels ($\beta = -0.0075$) were lower in older patients, whereas a higher expression level of *APOBEC3H* ($\beta = 0.0070$) was associated with increasing age.

Fig. 1 Effects of *cis*-eQTLs on the expression of 15 target genes. Shown are transcript levels according to the genotype of the top-ranking SNP for each eQTL. Genes shown in panels (A–O) are: *ABO*, *ANPEP*, *AP2A2*, *APOBEC3D*, *APOBEC3G*, *BSG*, *CLEC4G*, *DDX58*, *DPP4*, *FURIN*, *FYCO1*, *RAB14*, *SERINC3*, *TRIM5*, and *ZCRB1*. Expression levels are log₂-transformed probe intensity values. Genotypes are expressed as integers (0, 1, or 2) according to the number of minor alleles at each SNP. Parentheses indicate the number of the individuals carrying the indicated genotype. The line within each box represents the median; upper and lower edges of each box are 75th and 25th percentiles, respectively; top and bottom whiskers indicate the largest and smallest value within 1.5 times the interquartile range above the upper quartile and below the lower quartile, respectively; circles denote outliers (extreme values, >1.5 times the interquartile range). **P* < 0.05, ***P* < 0.01, ****P* < 0.001, ANOVA followed by Tukey's test for multiple comparisons.



Germline variations associated with gene expression

A genome-wide eQTL analysis was carried out for the 60 genes, considering 736,210 SNPs. This analysis identified 125 *cis*-eQTLs (FDR < 0.05) associated with the expression levels of 15 genes (Supplementary Table 3). *RAB14* gene

had the most *cis*-eQTLs (*n* = 50) and the top significant association (FDR = 7.9×10^{-53} , at rs56335605). Four transcripts (i.e., *AP2A2*, *DDX58*, *DPP4*, and *FURIN*) had only one *cis*-eQTL each, all with FDR > 0.01. No *trans*-eQTLs were found.

Figure 1 reports the top-ranking eQTL for each of the 15 genes showing genetic regulation of their expression in lung

tissue. For six genes (i.e., *ABO*, *APOBEC3D*, *BSG*, *CLEC4G*, *DPP4*, and *FURIN*), transcript levels increased with an increasing number of minor alleles of their eQTL SNP (i.e., rs8176722, rs139331, rs117582572, rs115293707, rs13390563, and rs4932179, respectively). The other nine genes (i.e., *ANPEP*, *AP2A2*, *APOBEC3G*, *DDX58I*, *FYCO1*, *RAB14*, *SERINC3*, *TRIM5*, and *ZCRB1*) showed lower transcript levels with an increasing number of the minor allele of rs11635469, rs11605303, rs8177832, rs73479410, rs936939, rs56335605, rs36121075, rs59814799, rs1563419, respectively.

Frequency, location, tissue specificity, and possible roles of eQTLs

To study these 15 *cis*-eQTLs, we considered the top-ranking SNP as representative of each locus. Table 1 shows the location of these SNPs with respect to their target gene, their MAF in different populations and in our series, and their predicted genetic functions. Five SNPs were intronic and eight mapped downstream of their target gene (from 0.545 to 587 kbp after the gene end). One SNP, rs56335605 (the most significant eQTL associated with *RAB14* levels), is located in the 3'UTR of its target gene. All but two of the 15 SNPs (i.e., rs8177832 and rs117582572) had MAF > 5% in our series, but MAFs varied widely in all populations examined.

Functional predictions were made for four of the top-ranking eQTL SNPs (Table 1 and Supplementary Table 4). SNPs rs11635469 (in *ANPEP*) and rs11605303 (in *AP2A2*) were predicted to affect transcription factor binding sites. SNP rs8177832, on the other hand, is a missense variant of *APOBEC3G* gene located in an exonic splicing enhancer; it was predicted to be a benign non-synonymous SNP (according to Polyphen) and to affect splicing activity. Finally, SNP rs56335605 (in *RAB14*) was predicted (by miRanda) to slightly alter the binding of some miRNAs. In particular, the presence of the minor allele of this SNP (A) was predicted to create a binding site for hsa-miR-1299.

Three of the 15 *cis*-eQTLs (in *ABO*, *ANPEP*, and *FYCO1* genes) have already been reported to be lung eQTLs in GTEx database (Table 1). The effects of genotype at these three SNPs in our study confirm those reported, with higher levels of *ABO* observed in individuals with a greater number of minor alleles and lower levels of *ANPEP* and *FYCO1* associated with an increasing number of minor alleles. The *APA2A* eQTL, instead, has been reported in ten tissues other than lung (e.g., brain, esophagus, colon, thyroid; not shown) with the same directions of the effect observed here. The *APOBEC3D* eQTL has also been reported in GTEx database, in a dozen tissues but not in lung, and with an opposite effect from what we found. SNP rs8177832 has been reported to be an *APOBEC3G* eQTL only in thyroid;

Table 1 Location, frequency, and predicted function of the best SNP at each of the 15 significant eQTLs.

eQTL SNP	eQTL target gene	Location (distance, kbp) ^a	Minor allele frequency (%) ^c							Functional prediction ^b	GTEx eQTL ^d	
			This study	AFR	AMR	EAS	SAS	EUR	Global		NES	P value
rs8176722	<i>ABO</i>	Intronic	9.1	14	6.5	19	24	11	15	None	0.55	4.9 × 10 ⁻³⁰
rs11635469	<i>ANPEP</i>	Intronic ^e	44	51	36	18	47	45	40	TFBS	-0.50	9.0 × 10 ⁻⁵⁴
rs11605303	<i>AP2A2</i>	Downstream (25.8)	8.5	8.3	6.1	0.1	1.4	8.2	5.0	TFBS	NA	NA
rs139331	<i>APOBEC3D</i>	Downstream (75.1)	32	32	24	25	39	34	31	None	NA	NA
rs8177832	<i>APOBEC3G</i>	Missense	4.2	43	7.5	7.2	0.8	3.0	15	Splicing	NA	NA
rs117582572	<i>BSG</i>	Downstream (9.14)	3.4	2.4	2.2	0.0	0.8	2.9	1.7	-	NA	NA
rs115293707	<i>CLEC4G</i>	Intronic	6.7	7.7	2.5	0.0	2.0	7.7	4.3	-	NA	NA
rs73479410	<i>DDX58</i>	Intronic	23	24	13	5.5	33	19	19	-	NA	NA
rs13390563	<i>DPP4</i>	Downstream (587)	12	57	10	34	15	13	29	None	NA	NA
rs4932179	<i>FURIN</i>	Downstream (4.22)	40	10	57	78	55	38	45	None	NA	NA
rs936939	<i>FYCO1</i>	Intronic	25	16	17	30	15	21	20	None	-0.18	5.1 × 10 ⁻⁹
rs56335605	<i>RAB14</i>	3'UTR	16	5	20	17	16	14	14	miRNA	NA	NA
rs36121075	<i>SERINC3</i>	Downstream (6.85)	14	7.3	6.8	0.1	10	13	7.5	None	NA	NA
rs59814799	<i>TRIM5</i>	Downstream (0.545)	13	10	8.8	0.0	1.5	13	6.7	None	NA	NA
rs1563419	<i>ZCRB1</i>	Downstream (19.2)	9.4	12	8.5	16	8.0	6.5	10	None	NA	NA

^aWith respect to target gene (UTR, untranslated region).

^bAs reported in SNPInfo Web Server (TFBS, transcription factor binding site; dashes indicate that the variant was not found by the tool); see Supplementary Table 4 for further details.

^cAFR Africans, AMR Americans, EAS East Asians, EUR Europeans, SAS South Asians (1000 Genomes Project Phase 3 [48]).

^dIn lung tissue (same SNP-target gene pair); NES normalized effect size, NA not available.

^eAnnotated in dbSNP as an upstream variant of *ANPEP* gene.

however, the second most significant eQTL SNP of this gene (i.e., the intronic rs17537581) has been reported as a lung eQTL with the same effects of genotype as in our study: lower levels of *APOBEC3G* associated with an increasing number of minor alleles (GTEx data: rs17537581 normalized effect size, -0.30 , $P = 2.2 \times 10^{-12}$). SNP rs117582572 has been reported to be an *BSG* eQTL only in esophageal mucosa. The rs115293707 variant has been not yet reported as an eQTL of *CLEC4G* gene in lung tissue, but it has in 14 other tissues. The best eQTL identified in our series (i.e., the one involving *RAB14* gene) has not been previously identified in lung; however, it has been found in the esophagus, but showing an opposite effect. On the contrary, the rs59814799 variant has already been reported as a *TRIM5* eQTL in tibial nerve, skeletal muscle, and adipose tissue, but not in lung; in all the tissues, the highest levels of *TRIM5* associated with the lowest number of minor alleles of rs59814799. No *DDX58* eQTLs have been previously identified in lung, and no associations between rs13390563 and *DPP4* (or any other transcript) levels, in any tissue, have been reported, according to GTEx database. Also, rs36121075 and rs1563419 had not been previously associated with *SERINC3* and *ZCRB1* levels, respectively.

Discussion

This study investigated the regulation of expression, in noninvolved lung tissue of surgically treated lung adenocarcinoma patients, of 60 genes selected for possible involvement in SARS-CoV-2 cell entry or in the host innate immune response. First, we did not observe significant differences in the expression of these genes between sexes, and age affected the expression levels of only two genes. From the eQTL analysis, genome-wide significant associations between 125 SNPs and the expression levels of 15 genes were identified.

Our lung gene expression analysis found that *APOBEC3H* and *PARP12* levels were significantly associated with age, with *APOBEC3H* levels higher and *PARP12* levels lower in older individuals. *APOBEC3H*, which encodes a polynucleotide cytosine deaminase, has several haplotypes and isoforms showing different antiviral activities [37]. Our results regard all *APOBEC3H* isoforms, as our microarray did not distinguish among them; whether the expression of different isoforms varies with age should be investigated, especially in COVID-19 patients, to understand if *APOBEC3H* isoforms affect prognosis in older patients. *PARP12* codes for an ADP-ribosyltransferase that has been suggested to modify Zika virus proteins, leading to their degradation and thereby inhibiting viral replication [38]. Although it is not yet known if this ADP-ribosyltransferase is involved in SARS-CoV-2 restriction,

if the herein observed correlation of *PARP12* levels with age are confirmed in the lung of COVID-19 patients, low *PARP12* expression in elders might explain their worse outcome.

The main result of this study is the identification of *cis*-eQTLs in lung tissue for 15 putative SARS-CoV-2 infection-associated genes. These eQTLs document the genetic control of expression of these genes in lung tissue. The genetic variants we found result in different levels of mRNA in different individuals, which is expected to result in different levels of the encoded proteins. Thus, individuals with different genotypes at these eQTL SNPs may be differently susceptible to SARS-CoV-2 infection. Although *ACE2* and *TMPRSS2* are the best candidates for differences in COVID-19 susceptibility, since they are directly involved in SARS-CoV-2 cell entry [17, 18], we did not find a significant association between their mRNA levels and any genetic variant. This result implies that their expression in noninvolved lung tissue is either not genetically regulated or subject to complex regulation that our study was unable to detect. Although the statistical power of our series was sufficient to detect relatively strong effects of single polymorphisms, it was underpowered to detect SNP interactions. In the lung, *TMPRSS2* has been observed to be co-expressed with *ACE2* in type II pneumocytes [39]. Since we did not observe a genetic control of *ACE2* in the lung, it is not surprising that even the expression of *TMPRSS2* in this tissue is not subject to strict genetic control.

We observed the genetic regulation of *BSG*, which encodes basigin (also known as CD147). This protein has been suggested to be an alternative receptor for SARS-CoV-2 adsorption and entry into human cells [20]. Therefore, individuals with different genotypes at the identified SNPs may express, in the lung, different levels of this putative SARS-CoV-2 receptor, resulting in different susceptibilities to the infection. We also found the genetic regulation of four genes (*ANPEP*, *CLEC4G*, *DPP4*, and *FURIN*) known to be involved in cell entry by other coronaviruses. These findings for *ANPEP* and *DPP4* conflict with a recent in vitro study that did not find a role for them in SARS-CoV-2 cell entry [40]. These findings are concordant, however, with another study that demonstrated that *FURIN* encodes a protease that cleaves the spike protein, in an essential step for lung cell infection by SARS-CoV-2 virus [41]. eQTLs were also found in genes required for viral genome replication (*ZCRB1*) and for membrane/vesicle trafficking (*AP2A2* and *RAB14*), and in genes encoding restriction factors (*APOBEC3D*, *APOBEC3G*, *DDX58*, *SERINC3*, and *TRIM5*).

The main limitation of this study is that it did not use samples from COVID-19 patients, nor for ethical reasons did it use samples of lung tissue from healthy persons. The samples were apparently normal lung tissue resected from

lung adenocarcinoma patients, and taken as far as possible from the tumor site. Although we cannot state that these samples are absolutely normal, the cellular composition (based on transcriptome profiles) was enriched in cell types normally present in lung tissue. This finding suggests that our results are not confounded by a heavy presence of infiltrating immune cells.

Some findings of our study are pertinent to the study of innate immune responses to viruses other than SARS-CoV-2. For instance, we identified eQTLs for two members of the APOBEC3 family (*APOBEC3D* and *APOBEC3G*), which encodes enzymes that interfere with the replication of RNA viruses, including some coronaviruses [42–44]. Therefore, individuals expressing different levels of these enzymes could respond differently to infection and develop more or less severe forms of other viral diseases for which a role of APOBEC3 enzymes in host responses has already been identified, e.g., HIV-1, HBV, HCV, and HCoV-NL63 [43, 45].

Finally, we identified *cis*-eQTLs in two COVID-19 susceptibility loci [16]. These eQTLs affect transcript levels of *ABO* at chr. 9q34.2 and *FYCO1* at chr. 3p21.31. Our data confirm previous reports on these eQTLs in lung, recorded in the GTEx database, and therefore support a role for these loci in susceptibility to severe COVID-19. Evidence that these eQTLs result in different abundances of the encoded proteins in lung tissue is required to confirm this role.

Studies on the role of host genetics are currently ongoing by several groups and international consortia (e.g., COVID-19 Host Genetics Initiative, <https://www.covid19hg.org/>, COVID Human Genetic Effort, www.covidhg.com). Recently other loci and variants have been found to associate with COVID-19 susceptibility or severity [46, 47]. In our opinion, a genome-wide approach, considering both coding and regulatory variants, is best for investigating the genetic basis of SARS-CoV-2 infection susceptibility and of interindividual differences in COVID-19 prognosis. Additionally, functional experiments are needed to understand the molecular mechanisms underlying the already identified associations between COVID-19 phenotypes and genetic variants.

Acknowledgements The authors acknowledge the contributions of Valerie Matarese, Ph.D., who provided scientific editing.

Funding This work was supported in part by a grant from the Fondazione Regionale Ricerca Biomedica (FRRB) of the Lombardy Region, Italy, to TAD.

Compliance with ethical standards

Conflict of interest All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: TAD had financial support (grant) from Fondazione Regionale per la

Ricerca Biomedica (Milan, Italy). The other authors had no support from any organization for the submitted work. All authors had no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval This case series is part of a broader research program on lung adenocarcinoma, for which a biobank of lung tissue and extracted nucleic acids is maintained at Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (previously described in [31]). Methods for tissue collection, storage and nucleic acid extraction, and information about ethical approval and patients' informed consent have already been reported [31].

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497–506.
- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708–20.
- Task force COVID-19 del Dipartimento Malattie Infettive e Servizio di Informatica, Istituto Superiore di Sanità. *Epidemia COVID-19, Aggiornamento nazionale: 2 dicembre 2020*. 2020; Available at: https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_2-dicembre-2020.pdf. Accessed 6 Dec 2020.
- Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J Infect*. 2020;81:e16–e25.
- Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. *Aging*. 2020;12:6049–57.
- Becerra-Flores M, Cardozo T. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *Int J Clin Pract*. 2020;74:e13525.
- Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet*. 2020;65:1075–82.
- Nogales A, DeDiego L, Host M. Single nucleotide polymorphisms modulating influenza a virus disease in humans. *Pathogens*. 2019;8:168. <https://doi.org/10.3390/pathogens8040168>.
- Allen EK, Randolph AG, Bhangale T, Dogra P, Ohlson M, Oshansky CM, et al. SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nat Med*. 2017;23:975–83.
- Cheng Z, Zhou J, To KK, Chu H, Li C, Wang D, et al. Identification of TMPRSS2 as a susceptibility gene for severe 2009 pandemic A(H1N1) influenza and A(H7N9) influenza. *J Infect Dis*. 2015;212:1214–21.
- Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature*. 2012;484:519–23.
- Tu X, Chong WP, Zhai Y, Zhang H, Zhang F, Wang S, et al. Functional polymorphisms of the CCL2 and MBL genes cumulatively increase susceptibility to severe acute respiratory syndrome coronavirus infection. *J Infect*. 2015;71:101–9.
- Zhu X, Wang Y, Zhang H, Liu X, Chen T, Yang R, et al. Genetic variation of the human alpha-2-Heremans-Schmid glycoprotein

- (AHS) gene associated with the risk of SARS-CoV infection. *PLoS ONE*. 2011;6:e23730.
14. Chan KY, Xu MS, Ching JC, So TM, Lai ST, Chu CM, et al. CD209 (DC-SIGN) -336A>G promoter polymorphism and severe acute respiratory syndrome in Hong Kong Chinese. *Hum Immunol*. 2010;71:702-7.
 15. Itoyama S, Keicho N, Quy T, Phi NC, Long HT, Ha LD, et al. ACE1 polymorphism and progression of SARS. *Biochem Biophys Res Commun*. 2004;323:1124-9.
 16. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med*. 2020;383:1522-34.
 17. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020;181:271-80.e8.
 18. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-3.
 19. Hussain M, Jabeen N, Raza F, Shabbir S, Baig AA, Amanullah A, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J Med Virol*. 2020;92:1580-6.
 20. Wang K, Chen W, Zhou Y, Lian J, Zhang Z, Du P, et al. SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. *bioRxiv*. 2020. <https://www.biorxiv.org/content/10.1101/2020.03.14.988345v1>.
 21. Singh M, Bansal V, Feschotte C. A single-cell RNA expression map of human coronavirus entry factors. *Cell Rep*. 2020;32:108175.
 22. Chemudupati M, Kenney AD, Bonifati S, Zani A, McMichael TM, Wu L, et al. From APOBEC to ZAP: diverse mechanisms used by cellular restriction factors to inhibit virus infections. *Biochim Biophys Acta Mol Cell Res*. 2019;1866:382-94.
 23. Bailey CC, Zhong G, Huang IC, Farzan M. IFITM-family proteins: the cell's first line of antiviral defense. *Annu Rev Virol*. 2014;1:261-83.
 24. Heraud-Farlow JE, Walkley CR. The role of RNA editing by ADAR1 in prevention of innate immune sensing of self-RNA. *J Mol Med*. 2016;94:1095-102.
 25. Grunewald ME, Chen Y, Kuny C, Maejima T, Lease R, Ferraris D, et al. The coronavirus macrodomain is required to prevent PARP-mediated inhibition of virus replication and enhancement of IFN expression. *PLoS Pathog*. 2019;15:e1007756.
 26. Zhou J, To KK, Dong H, Cheng ZS, Lau CC, Poon VK, et al. A functional variation in CD55 increases the severity of 2009 pandemic H1N1 influenza A virus infection. *J Infect Dis*. 2012;206:495-503.
 27. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*. 2013;368:20120362.
 28. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24:1547-8.
 29. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28:882-3.
 30. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18:220.
 31. Maspero D, Dassano A, Pintarelli G, Noci S, De Cecco L, Incarbone M, et al. Read-through transcripts in lung: germline genetic regulation and correlation with the expression of other genes. *Carcinogenesis*. 2020;41:918-26.
 32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-75.
 33. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc*. 2010;5:1564-73.
 34. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012;28:1353-8.
 35. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 2001;125:279-84.
 36. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37:W600-5.
 37. Harari A, Ooms M, Mulder LC, Simon V. Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. *J Virol*. 2009;83:295-303.
 38. Li L, Zhao H, Liu P, Li C, Quanquin N, Ji X, et al. PARP12 suppresses Zika virus infection through PARP-dependent degradation of NS1 and NS3 viral proteins. *Sci Signal*. 2018;11. <https://doi.org/10.1126/scisignal.aas9332>.
 39. Ziegler CGK, Allon SJ, Nyquist SK, Mbano IM, Miao VN, Tzouanas CN, et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell*. 2020;181:1016-35.e19.
 40. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B beta-coronaviruses. *Nat Microbiol*. 2020;5:562-9.
 41. Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell*. 2020;78:779-84.e5.
 42. Krisko JF, Begum N, Baker CE, Foster JL, Garcia JV. APOBEC3G and APOBEC3F act in concert to extinguish HIV-1 replication. *J Virol*. 2016;90:4681-95.
 43. Milewska A, Kindler E, Vkovski P, Zeglen S, Ochman M, Thiel V, et al. APOBEC3-mediated restriction of RNA virus replication. *Sci Rep*. 2018;8:5960-018-24448-2.
 44. Li J, Chen Y, Li M, Carpenter MA, McDougale RM, Luengas EM, et al. APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *J Mol Biol*. 2014;426:1296-307.
 45. Siriwardena SU, Chen K, Bhagwat AS. Functions and malfunctions of mammalian DNA-cytosine deaminases. *Chem Rev*. 2016;116:12688-710.
 46. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*. 2020;370:eabd4570. <https://doi.org/10.1126/science.abd4570>.
 47. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick A, Rawlik K, Parkinson N, et al. Genetic mechanisms of critical illness in Covid-19. *medRxiv*. 2020. <https://www.medrxiv.org/content/10.1101/2020.09.24.20200048v2>.
 48. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74.