Journal of Translational Medicine

RESEARCH

Open Access

Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2

Abul Bashar Mir Md. Khademul Islam[1*†], Md. Abdullah-Al-Kamran Khan[2†], Rasel Ahmed[3], Md. Sabbir Hossain[3], Shah Md. Tamim Kabir[3], Md. Shahidul Islam[3] and A. M. A. M. Zonaed Siddiki[4]

Abstract

Background: Although it is becoming evident that individual's immune system has a decisive influence on SARS-CoV-2 disease progression, pathogenesis is largely unknown. In this study, we aimed to profile the host transcriptome of COVID-19 patients from nasopharyngeal samples along with virus genomic features isolated from respective host, and a comparative analyses of differential host responses in various SARS-CoV-2 infection systems.

Results: Unique and rare missense mutations in 3C-like protease observed in all of our reported isolates. Functional enrichment analyses exhibited that the host induced responses are mediated by innate immunity, interferon, and cytokine stimulation. Surprisingly, induction of apoptosis, phagosome, antigen presentation, hypoxia response was lacking within these patients. Upregulation of immune and cytokine signaling genes such as *CCL4, TNFA, IL6, IL1A, CCL2, CXCL2, IFN,* and *CCR1* were observed in lungs. Lungs lacked the overexpression of ACE2 as suspected, however, high *ACE2* but low *DPP4* expression was observed in nasopharyngeal cells. Interestingly, directly or indirectly, viral proteins specially non-structural protein mediated overexpression of integrins such as *ITGAV, ITGA6, ITGB7, ITGB3, ITGA2B, ITGA5, ITGA6, ITGA9, ITGA4, ITGAE,* and *ITGA8* in lungs compared to nasopharyngeal samples suggesting the possible way of enhanced invasion. Furthermore, we found comparatively highly expressed transcription factors such as CBP, CEBP, NFAT, ATF3, GATA6, HDAC2, TCF12 which have pivotal roles in lung injury.

Conclusions: Even though this study incorporates a limited number of cases, our data will provide valuable insights in developing potential studies to elucidate the differential host responses on the viral pathogenesis in COVID-19, and incorporation of further data will enrich the search of an effective therapeutics.

Keywords: Host transcriptional response, COVID-19, SARS-CoV-2, Genome variations, Immune response, Integrins

Background

Since the declaration of COVID-19 pandemic on 11 March, this Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) mediated infection has spread ~ 213 countries and territories [1]. Approximately, 15 million individuals across the globe have fallen victim to this virus and the number is constantly increasing

*Correspondence: khademul@du.ac.bd
†Abul Bashar Mir Md. KhademulIslam and Md. Abdullah-Al-Kamran Khan contributed equally to this work.
[1] Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka 1000, Bangladesh
Full list of author information is available at the end of the article

at an alarming rate, as of the writing of this manuscript [1]. Though the initial fatality was as low as 3.5%, currently this value lies around ~ 6.66% [1] and it might be increased because of the withdrawal of earlier preventing measures taken throughout the world. Coronaviruses are not new to human civilization, as these viruses caused several earlier outbreaks during the past two decades. However, none of the earlier outbreaks spread as widely as the current ongoing pandemic. As the pandemic progresses, more researches on the molecular pathobiology of the COVID-19 are being rapidly carried out to search for effective therapeutic intervention.

Coronaviruses possess single-stranded RNA (positive sense) genomes lengthening approximately 30 Kb [2]. Amongst the coronaviruses, SARS-CoV-2 is a member of the betacoronaviruses having a ~ 29.9 Kb genome which contains 11 functional genes [3]. Though SARS-CoV-2 shows similar clinical characteristics as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and Middle East Respiratory Syndrome-related Coronavirus (MERS-CoV) viruses, it has only ~ 79% and ~ 50% genome sequence similarities with these viruses, respectively; whereas, the genome sequence of SARS-CoV-2 is ~ 90% identical to that of bat derived SARS-like coronavirus [4]. Moreover, several key genomic variances between SARS-CoV-2 and SARS-CoV such as- 380 different amino acid substitutions, ORF8a deletion, ORF8b elongation, and ORF3b truncation were also reported [2].

The clinical characteristics of the COVID-19 range from mild fever to severe lung injury [5]. Some of the commonly observed mild COVID-19 symptoms are fever, cough, and fatigue; however, complications such as- myalgia, shortness of breath, headache, diarrhea, and sore throat were also reported [6]. Furthermore, severely affected patients had exhibited respiratory complications like moderate to severe pneumonia, acute respiratory distress syndrome (ARDS), sepsis, acute lung injury (ALI), and multiple organ dysfunction (MOD) [7]. Primarily, the lungs of the COVID-19 patients are affected [8]; however, failures of other functional systems, namely cardiovascular system, and nervous system were also reported [9, 10].

Several features of the SARS-CoV-2 infection made it more complicated for effective clinical management. From the earlier studies, the incubation period of SARS-CoV-2 was reported to be around 4–5 days, however, some recent studies suggested a prolonged incubation period of 8–27 days [11]. Additionally, several cases of viral latency within the host [12], and the recurrent presence of SARS-CoV-2 in clinically recovered patients were also recorded [13, 14]. However, the detailed molecular mechanism behind these phenomena is still elusive.

In COVID-19, an increased level of infection-associated pro-inflammatory cytokines were recorded [15], which thereby supports the term "Cytokine storm", that was frequently used to describe the SARS-CoV and MERS-CoV disease pathobiology [16]. This phenomenon causes the hyperactivation and recruitment of the inflammatory cells within the lungs and results in the acute lung injury of the infected patients [17]. However, this illustrates one putative molecular mechanism of COVID-19, there are many other immune regulators and host genetic/epigenetic factors which can also play significant contribution towards the disease manifestation [18, 19]. This multifaceted regulation was also reported previously for other different coronavirus infections [20]. Host–pathogen interactions in different coronavirus infections can function as a double-edged sword, as these could be beneficial not only to the hosts but also the viruses [20]. Similar host-virus tug-of-war can also occur in COVID-19 which might be contributing towards the overcomplicated disease outcomes [21].

Collectively, more than 1.7 million (almost 9% of the total infections around the globe) people have been diagnosed with COVID-19 in the South-Asian region and the number is still increasing devastatingly [1]. Recently, it has been speculated that South-Asian people might be possessing a genomic region acting as the risk factor for COVID-19 [22]. Moreover, another study suggested some genomic variations in several Indian SARS-CoV-2 isolates that might be involved in the COVID-19 pathogenesis in Indian patients [23]. However, any data suggesting the COVID-19 patients' transcriptomic responses from this part of the globe are yet to be reported.

SARS-CoV-2 follows a highly variable course and it is becoming more evident that individual's immune system has a decisive influence on the progression of the disease [24]. However, the detailed underlying molecular mechanisms of the SARS-CoV-2 mediate disease pathogenesis are largely unknown. Even previously conducted studies using patient samples, animal models, and cell lines to explain the pathobiology of COVID-19 [24–26] lack a detailed comparison of the host transcriptional responses between different infection models as well as the different sites of the respiratory system that might provide valuable insights on the COVID-19 pathogenesis and disease severity. In this present study, we sought to discuss the host transcriptional responses observed in naspharyangeal cells of COVID-19 patients. This trscriptional profile report is first such kind from South Asian region. Additionally, we reported the genome variations observed in the four SARS-CoV-2 isolates obtained from these patients. Finally, we illuminated the differences in host transcriptional responses in different COVID-19 infection models and further pursued to discover the putative effects of these altered responses (Fig. 1).

Results

Our sequenced SARS-CoV-2 isolates showed a divergent variation pattern compared to the other worldwide isolates

We sought to find out the genome variations within the four SARS-CoV-2 isolates that we sequenced, and pursued the deviation of these genomes compared to the other isolates from this country. To accomplish these goals, we first identified and annotated the genome variations observed within our sequenced isolates. Then we produced informative statistics from these observed variations and compared the prevalence of those with the other isolates of Bangladesh and the rest of the world.

We mapped the RNA-seq reads of each of the samples and checked their distribution athwart the entire reference genome of SARS-CoV-2 (Fig. 2a). High coverages and read evidence were observed for all the isolates across the whole genome of the SARS-CoV-2 (Fig. 2a). This suggests that the sequenced genomes of these isolates are of high coverage and no such region is observed without the mapped reads.

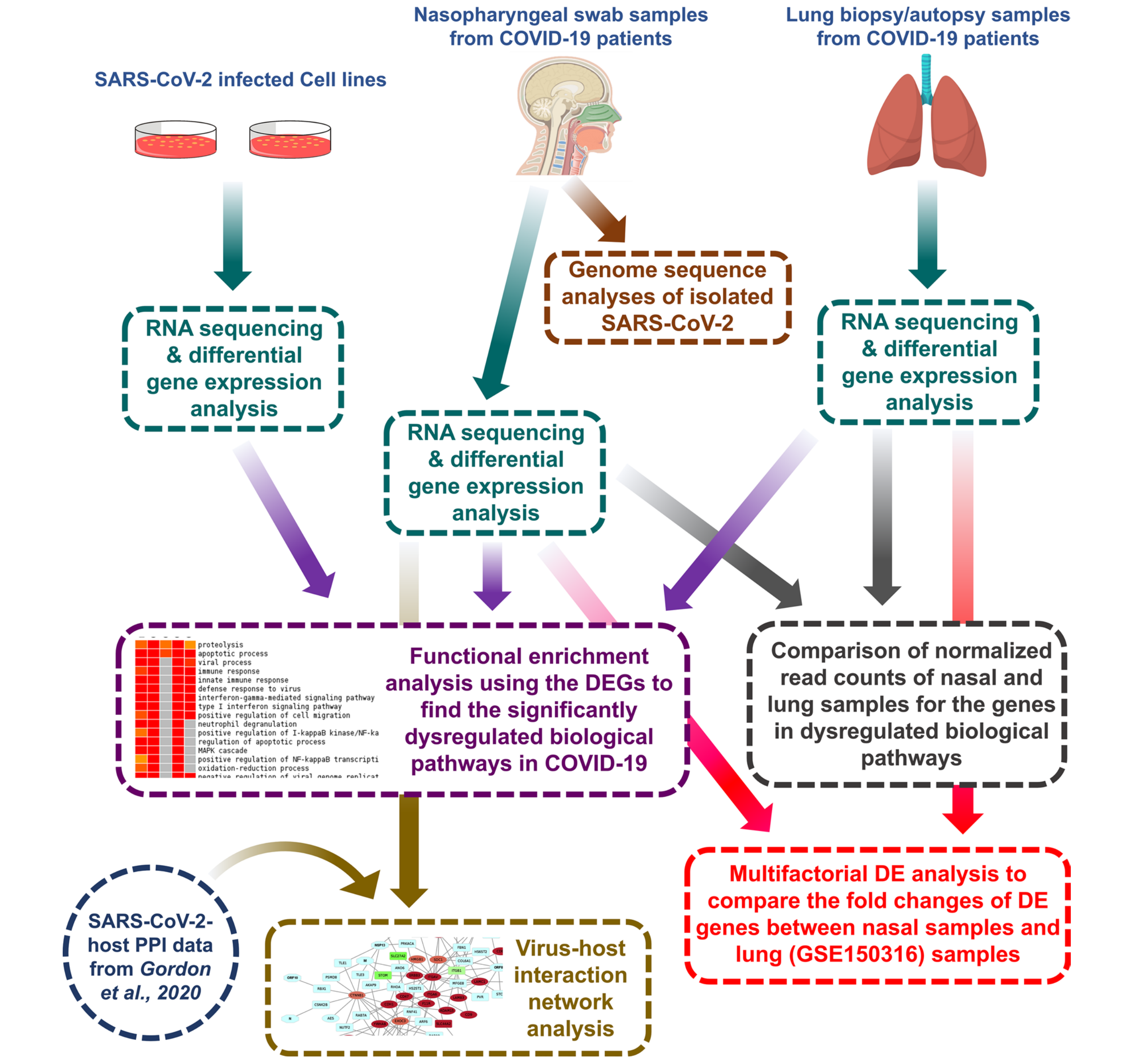

Nasopharyngeal swab samples from COVID-19 patients
Lung biopsy/autopsy samples from COVID-19 patients
SARS-CoV-2 infected Cell lines
Genome sequence analyses of isolated SARS-CoV-2
RNA sequencing & differential gene expression analysis
RNA sequencing & differential gene expression analysis
RNA sequencing & differential gene expression analysis
proteolysis
apoptotic process
viral process
immune response
innate immune response
defense response to virus
interferon-gamma-mediated signaling pathway
type I interferon signaling pathway
positive regulation of cell migration
neutrophil degranulation
positive regulation of I-kappaB kinase/NF-ka
regulation of apoptotic process
MAPK cascade
positive regulation of NF-kappaB transcripti
oxidation-reduction process
negative regulation of viral genome replicat
Functional enrichment analysis using the DEGs to find the significantly dysregulated biological pathways in COVID-19
Comparison of normalized read counts of nasal and lung samples for the genes in dysregulated biological pathways
Multifactorial DE analysis to compare the fold changes of DE genes between nasal samples and lung (GSE150316) samples
SARS-CoV-2-host PPI data from *Gordon et al.*, 2020
Virus-host interaction network analysis

Fig. 1 A brief workflow of the whole study

(See figure on next page.)

Fig. 2 Genomic information of the sequenced SARS-CoV-2 isolates. **a** Genome coverage normalized density map for the four sequenced SARS-CoV-2 isolates. **b** Pie-chart illustrating the different types of variations found within these four isolates. **c** Genome location-wise representation of the mutations and their associated frequency. **d** Isolate-wise variation information. **e** Gene-wise amount and type of mutations. **f** Annotated impacts of the different mutations (only those are shown which have frequencies more than 1). **g** Frequencies of selected unique mutations observed in these isolates

We detected sixty different types of variations within these four analyzed SARS-CoV-2 isolates (Table 1). All the four different types of sequence variations were spotted in these isolates, however, single nucleotide polymorphisms (SNPs) were the most prominent (Fig. 2b). Among these variations, twelve variations were found in more than one isolate, whereas rest forty-eight variation occurred in only one isolate (Table 1, Fig. 2c). Among the isolates, isolate S3 contained the lowest number of variations, whereas isolate S4 has the highest number of variations (Fig. 2d). Many concepts are there correlating the probable roles of variations with the COVID-19 disease severity [27, 28]. We did not observe any such variations within the spike region of our reported isolates; however, we recorded an unusual amount of 3′-UTR and 5′-UTR variations within these four isolates (Fig. 2c, d). Surprisingly, out of all these variations, we found only one downstream gene variation on the 3′-UTRs of all the four isolates; this variation can potentially impact the regulation of the ORF10 gene (Fig. 2d). Most of the nucleic acid mutations were located on the 3′-UTR of the isolates, whereas the ORF1ab gene contained most of the amino acid mutations (Fig. 2e).

No highly severe mutation was identified amongst these variations, but we found nine moderately impacting, seven low impacting, and forty-seven modifier variations within these isolates (Fig. 2f, Additional file 1). As of 8th July, thirty-eight out of the sixty variations within our sequenced isolates were completely absent in all other SARS-CoV-2 isolates (Table 1). Strikingly, we observed that variation 10,329: A>G is present within three of our sequenced isolates, only one other Bangladeshi and one other USA isolate contain this variation (Fig. 2g). This variation is located within the 3C-like protease of SARS-CoV-2. Previously, the potential implication of the mutations of this protein was reported to alter its overall structure and functionality [29–31] in SARS-CoV. The only one deceased patient did not have this mutation in our samples. Also, few of our reported variations like 25,505: A>T and 29,392: G>T are not highly prevalent globally (Fig. 2g).

Exploring the Nextstrain portal [32], we noticed that our analyzed SARS-CoV-2 sequences are closely placed to the Saudi-Arabian isolates (Additional file 2: Figure S1A); although, most of the other isolates of this country were placed in the major European clusters (data not shown). Furthermore, these isolates analyzed in this study are distinctly placed in our constructed Neighbor-Joining phylogenetic tree (Additional file 2: Figure S1B), this also supports the differences between these isolates and other SARS-CoV-2 isolates of this country which might have been originated from the European nations. As a large number of people from Bangladesh recently immigrated to Middle-East (particularly Saudi Arabia) for work [33]; those immigrant people returning from the Middle-East during this pandemic might have brought these isolates into Bangladesh.

Stimulated antiviral immune responses are detected in the nasopharyngeal samples of COVID-19 patients

Our analyzed patients exhibited the commonly observed sign and symptoms of COVID-19 such as mild fever, sore throat, coughing, bodyache, fatigue, and dysosmia (Additional file 3). Patients were hospitalized but no intensive clinical interventions such as ICU support or ventilation support were needed. Male to female ratio of the patients were 1:1. The median age of the patients were ~ 45 years, only one patient was around 85 years old. This oldest patient had some additional clinical features such as pre-existing asthma and diarrhea. All the patients recovered within one month of the initial diagnosis except patient S9, who died after COVID-19 infection.

Though initial researches suggested the potential implication of viral variations on the COVID-19 disease severity, one recent study indicated otherwise; Several host factors such as abnormal immune responses, and cytokine signaling might be influencing the overall disease outcomes more prominently compared to the viral mutations [34]. Moreover, several data surmised that ethnicity might be a pivotal risk factor of being susceptible to COVID-19 [35].

In this context, we explored the transcriptome data obtained from the nasopharyngeal samples from COVID-19 patients to find out how these patients were responding against the invading SARS-CoV-2. We compared the RNA-seq data of these patients with some random normal individuals' nasopharyngeal RNA-seq data to find out the differentially expressed genes within our analyzed samples. We observed a roughly constant standard deviation for the normalized reads suggesting a lesser amount of variation occurred during the normalization (Fig. 3a). Furthermore, we performed sample

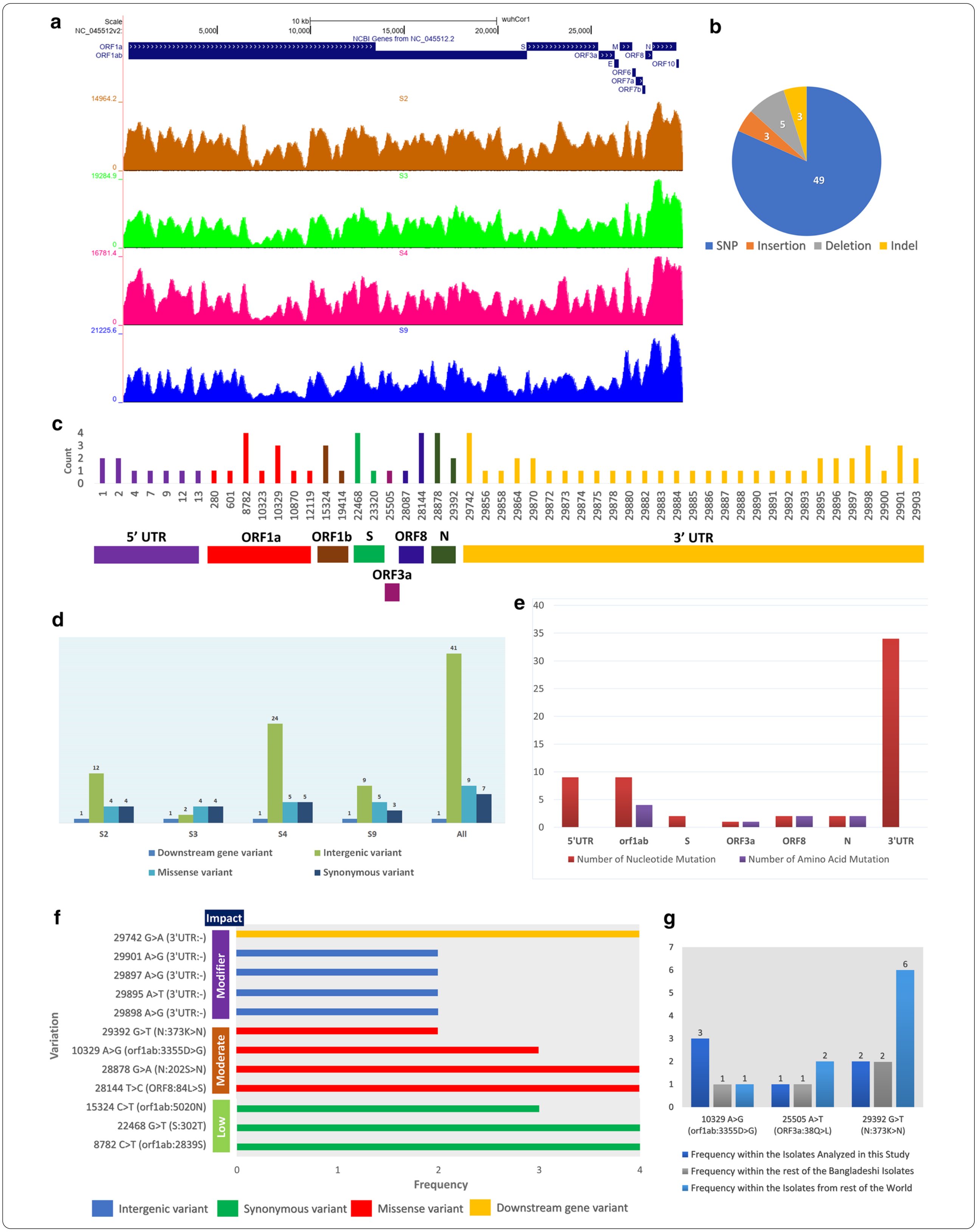

a
Scale
NC_045512v2:
10 kb
wuhCor1
5,000
10,000
15,000
20,000
25,000
NCBI Genes from NC_045512.2
ORF1a
ORF1ab
S
M
N
ORF3a
ORF8
E
ORF10
ORF6
ORF7a
ORF7b
14964.2
S2
0
19284.9
S3
0
16781.4
S4
0
21225.6
S9
0
b
49
3
5
3
SNP
Insertion
Deletion
Indel
c
Count
4
3
2
1
0
1
2
4
7
9
12
13
280
601
8782
10323
10329
10870
12119
15324
19414
22468
23320
25505
28087
28144
28878
29392
29742
29856
29858
29864
29870
29872
29873
29874
29875
29878
29880
29882
29883
29884
29885
29886
29887
29888
29890
29891
29892
29893
29895
29896
29897
29898
29900
29901
29903
5' UTR
ORF1a
ORF1b
S
ORF8
N
3' UTR
ORF3a
d
1
12
4
4
1
2
4
4
1
24
5
5
1
9
5
3
1
41
9
7
S2
S3
S4
S9
All
Downstream gene variant
Intergenic variant
Missense variant
Synonymous variant
e
40
35
30
25
20
15
10
5
0
5'UTR
orf1ab
S
ORF3a
ORF8
N
3'UTR
Number of Nucleotide Mutation
Number of Amino Acid Mutation
f
Impact
Variation
29742 G>A (3'UTR:-)
29901 A>G (3'UTR:-)
29897 A>G (3'UTR:-)
29895 A>T (3'UTR:-)
29898 A>G (3'UTR:-)
29392 G>T (N:373K>N)
10329 A>G (orf1ab:3355D>G)
28878 G>A (N:202S>N)
28144 T>C (ORF8:84L>S)
15324 C>T (orf1ab:5020N)
22468 G>T (S:302T)
8782 C>T (orf1ab:2839S)
Modifier
Moderate
Low
0
1
2
3
4
Frequency
Intergenic variant
Synonymous variant
Missense variant
Downstream gene variant
g
7
6
5
4
3
2
1
0
3
1
1
1
1
2
2
2
6
10329 A>G
(orf1ab:3355D>G)
25505 A>T
(ORF3a:38Q>L)
29392 G>T
(N:373K>N)
Frequency within the Isolates Analyzed in this Study
Frequency within the rest of the Bangladeshi Isolates
Frequency within the Isolates from rest of the World

Table 1 Observed variations within the four SARS-CoV-2 isolates reported in this study

Genomic position: Variation	Variation Type	Associated genomic region	Protein: amino acid change	Frequency in the four isolates used in this study	Frequency in the other Bangladeshi isolates	Frequency in the isolates from rest of the world
1: ATTAAAGGTTTA>-	Intergenic variant	5′UTR	–	1	–	–
1: ATTAAAGGTTTATA>-	Intergenic variant	5′UTR	–	1	–	–
2: T>TA	Intergenic variant	5′UTR	–	1	0	0
2: T>TTTCAAAGATCAAGTCA	Intergenic variant	5′UTR	–	1	0	0
4: A>T	Intergenic variant	5′UTR	–	1	0	58
7: G>C	Intergenic variant	5′UTR	–	1	0	16
9: T>TTTTCGC	Intergenic variant	5′UTR	–	1	0	0
12: A>T	Intergenic variant	5′UTR	–	1	0	22
13: T>C	Intergenic variant	5′UTR	–	1	0	36
280: C>T	Synonymous variant	orf1ab	5 V	1	0	5
601: C>T	Synonymous variant	orf1ab	112G	1	1	6
8782: C>T	Synonymous variant	orf1ab	2839S	4	1	3012
10,323: A>G	Missense variant	orf1ab	3353 K>R	1	5	154
10,329: A>G	Missense variant	orf1ab	3355D>G	3	1	1
10,870: G>T	Synonymous variant	orf1ab	3535L	1	0	27
12,119: C>T	Missense variant	orf1ab	3952P>S	1	0	8
15,324: C>T	Synonymous variant	orf1ab	5020 N	3	5	818
19,414: G>A	Missense variant	orf1ab	6384 V>I	1	0	0
22,468: G>T	Synonymous variant	S	302 T	4	1	99
23,320: C>T	Synonymous variant	S	586D	1	0	2
25,505: A>T	Missense variant	ORF3a	38Q>L	1	0	2
28,087: C>T	Missense variant	ORF8	65A>V	1	0	23
28,144: T>C	Missense variant	ORF8	84L>S	4	1	3050
28,878: G>A	Missense variant	N	202S>N	4	1	253
29,392: G>T	Missense variant	N	373 K>N	2	2	6
29,742: G>A	Downstream gene variant	3′UTR	–	4	1	21
29,856: T>A	Intergenic variant	3′UTR	–	1	0	6
29,858: T>A	Intergenic variant	3′UTR	–	1	0	5
29,864: GAATGACAAAAAAAAAAAAAAAAAAAAA>G	Intergenic variant	3′UTR	–	1	0	0
29,864: GAATGACAAAAAAAAAAAAAAAAAAAAAAA>T	Intergenic variant	3′UTR	–	1	0	0
29,870: CAAAAAAAAAAAAAAAAAAAAAAAAAA>C	Intergenic variant	3′UTR	–	1	1	–
29,870: C>G	Intergenic variant	3′UTR	–	1	0	3
29,872: A>T	Intergenic variant	3′UTR	–	1	0	12
29,873: A>C	Intergenic variant	3′UTR	–	1	0	3
29,874: A>G	Intergenic variant	3′UTR	–	1	0	12
29,875: A>G	Intergenic variant	3′UTR	–	1	1	5
29,878: A>T	Intergenic variant	3′UTR	–	1	0	3
29,880: A>G	Intergenic variant	3′UTR	–	1	2	5
29,882: A>G	Intergenic variant	3′UTR	–	1	0	13
29,883: A>T	Intergenic variant	3′UTR	–	1	0	8

Table 1 (continued)

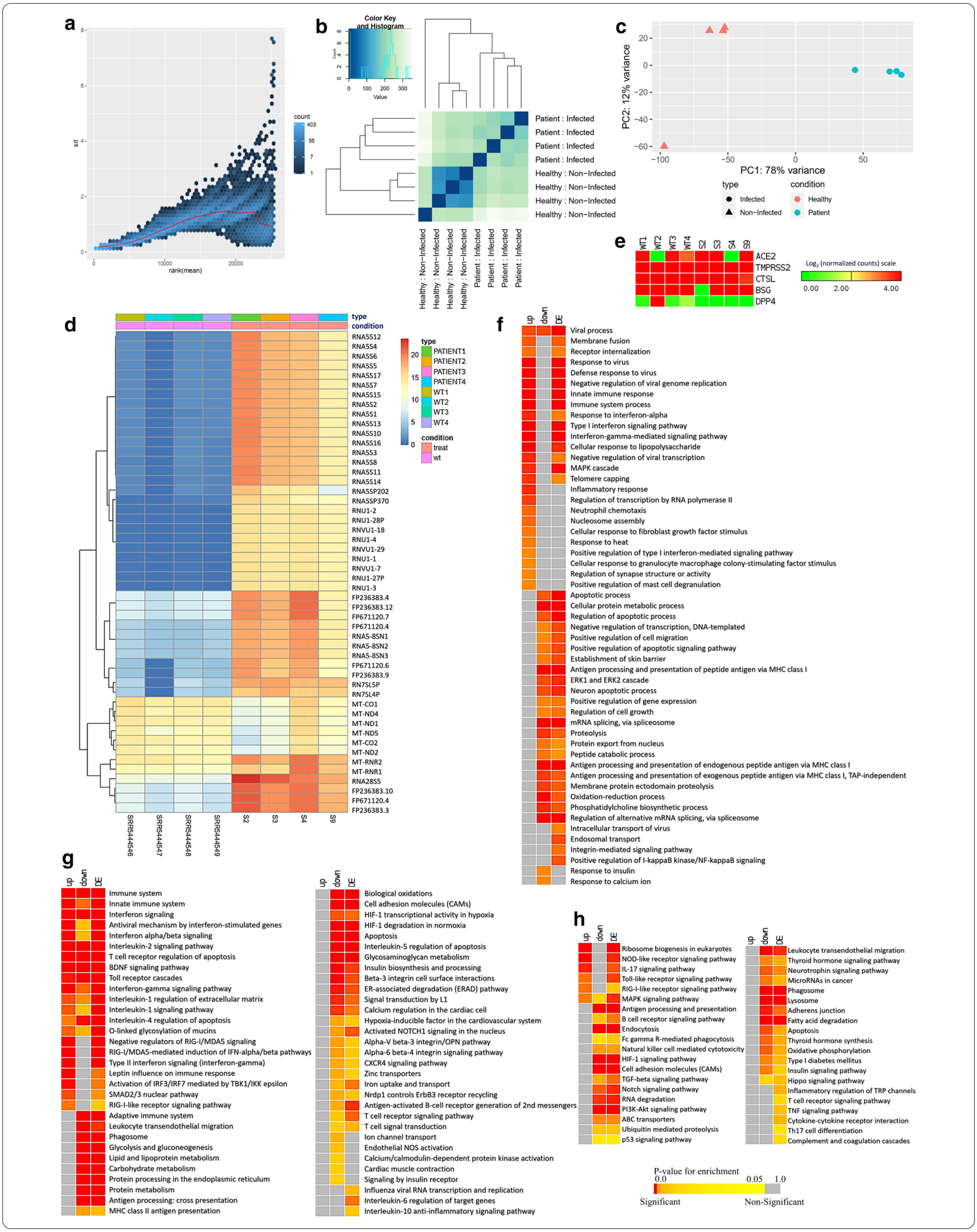
Genomic position: Variation	Variation Type	Associated genomic region	Protein: amino acid change	Frequency in the four isolates used in this study	Frequency in the other Bangladeshi isolates	Frequency in the isolates from rest of the world
29,884: A>C	Intergenic variant	3′UTR	–	1	0	10
29,885: A>G	Intergenic variant	3′UTR	–	1	1	11
29,886: A>T	Intergenic variant	3′UTR	–	1	1	5
29,887: A>G	Intergenic variant	3′UTR	–	1	0	11
29,888: A>T	Intergenic variant	3′UTR	–	1	1	4
29,890: A>G	Intergenic variant	3′UTR	–	1	1	6
29,891: A>G	Intergenic variant	3′UTR	–	1	0	15
29,892: A>G	Intergenic variant	3′UTR	–	1	0	13
29,893: A>G	Intergenic variant	3′UTR	–	1	0	15
29,895: A>T	Intergenic variant	3′UTR	–	2	0	6
29,896: A>G	Intergenic variant	3′UTR	–	1	0	7
29,896: A>C	Intergenic variant	3′UTR	–	1	0	3
29,897: A>G	Intergenic variant	3′UTR	–	2	0	4
29,898: A>G	Intergenic variant	3′UTR	–	2	0	6
29,898: A>T	Intergenic variant	3′UTR	–	1	0	5
29,900: A>G	Intergenic variant	3′UTR	–	1	0	10
29,901: AAA>A	Intergenic variant	3′UTR	–	1	0	–
29,901: A>G	Intergenic variant	3′UTR	–	2	0	5
29,903: A>GCCGTCGT	Intergenic variant	3′UTR	–	1	0	–
29,903: A>GCGTCGTGT	Intergenic variant	3′UTR	–	1	0	–

(See figure on next page.)

Fig. 3 Differential gene expression analysis of the studied nasophryngeal samples of COVID-19 patients. **a** Variance plot. This plots the standard deviation of the transformed data, across samples, against the mean, using the variance stabilizing transformation. The vertical axis in the plots is the square root of the variance over all samples. **b** Sample to sample distance plot. A heatmap of distance matrix providing an overview of similarities and dissimilarities between samples. Clustering is based on the distances between the rows/columns of the distance matrix. **c** Principal component analysis plot. Samples are in the 2D plane spanned by their first two principal components. **d** Clustered heatmap of the $\log_2$ converted normalized count matrix RNA-seq reads, top 50 genes, of nasopharyngeal samples. **e** Normalized Log_2 read counts of the genes encoding SARS-CoV-2 receptor and entry associated proteins. Enrichment analysis and comparison between deregulated genes and the genes of some selected processes in SARS-CoV-2 infected nasopharyngeal samples and SARS-CoV-2 infected lung biopsy samples using **f** GOBP module, **g** KEGG pathway, **h** Bioplanet pathway module. Selected significant terms are represented in heatmaps. Significance of enrichment in terms of the adjusted p-value (< 0.05) is represented in color-coded P-value scale for all heatmaps; Color towards red indicates higher significance and color towards yellow indicates less significance, while grey means non-significant. Normalized Log_2 converted read counts are considered as the expression values of the genes and represented in a color-coded scale; Color towards red indicating higher expression, while color towards green indicating little to no expression. Here, Up, down and DE denote Upregulated, Downregulated and Differentially expressed, respectively

clustering to assess the quality of our generated normalized RNA-seq data. No anomalies were observed in the sample to sample distance matrix (Fig. 3b) and principal component analysis (PCA) (Fig. 3c) while comparing our samples with the used healthy individuals' data. Moreover, the larger differences observed in the PCA plot (Fig. 3c) and clustered heatmap of the count matrix with the top 50 significant genes (Fig. 3d) suggest a significant transcriptomic response difference between our infected patients' data and the normal individuals' data. Likewise, the sample to sample distance plot suggested the similarities of samples of similar nature; the infected and healthy samples were clustered into two distinct groups (Fig. 3b).

Sungnak et al. described the significance of several viral entry associated host proteins in SARS-CoV-2 pathogenesis, namely ACE2, TMPRSS2, BSG, CTSL, DPP4 [36]. We also investigated the expression of the associated transcripts of these proteins within our


a
sd
rank(mean)
count
403
55
7
1
b
Color Key and Histogram
Count
Value
Patient : Infected
Patient : Infected
Patient : Infected
Patient : Infected
Healthy : Non−Infected
Healthy : Non−Infected
Healthy : Non−Infected
Healthy : Non−Infected
c
PC1: 78% variance
PC2: 12% variance
type
Infected
Non−Infected
condition
Healthy
Patient
e
WT1
WT2
WT3
WT4
S2
S3
S4
S9
ACE2
TMPRSS2
CTSL
BSG
DPP4
Log2 (normalized counts) scale
0.00
2.00
4.00
d
type
condition
PATIENT1
PATIENT2
PATIENT3
PATIENT4
WT1
WT2
WT3
WT4
treat
wt
RNA5S12
RNA5S4
RNA5S6
RNA5S5
RNA5S17
RNA5S7
RNA5S15
RNA5S2
RNA5S1
RNA5S13
RNA5S10
RNA5S16
RNA5S3
RNA5S8
RNA5S11
RNA5S14
RNA5SP202
RNA5SP370
RNU1-2
RNU1-28P
RNVU1-18
RNU1-4
RNVU1-29
RNU1-1
RNVU1-7
RNU1-27P
RNU1-3
FP236383.4
FP236383.12
FP671120.7
FP671120.4
RNA5-8SN1
RNA5-8SN2
RNA5-8SN3
FP671120.6
FP236383.9
RN7SL5P
RN7SL4P
MT-CO1
MT-ND4
MT-ND1
MT-ND5
MT-CO2
MT-ND2
MT-RNR2
MT-RNR1
RNA28S5
FP236383.10
FP671120.4
FP236383.3
SRR5444546
SRR5444547
SRR5444548
SRR5444549
S2
S3
S4
S9
f
up
down
DE
Viral process
Membrane fusion
Receptor internalization
Response to virus
Defense response to virus
Negative regulation of viral genome replication
Innate immune response
Immune system process
Response to interferon-alpha
Type I interferon signaling pathway
Interferon-gamma-mediated signaling pathway
Cellular response to lipopolysaccharide
Negative regulation of viral transcription
MAPK cascade
Telomere capping
Inflammatory response
Regulation of transcription by RNA polymerase II
Neutrophil chemotaxis
Nucleosome assembly
Cellular response to fibroblast growth factor stimulus
Response to heat
Positive regulation of type I interferon-mediated signaling pathway
Cellular response to granulocyte macrophage colony-stimulating factor stimulus
Regulation of synapse structure or activity
Positive regulation of mast cell degranulation
Apoptotic process
Cellular protein metabolic process
Regulation of apoptotic process
Negative regulation of transcription, DNA-templated
Positive regulation of cell migration
Positive regulation of apoptotic signaling pathway
Establishment of skin barrier
Antigen processing and presentation of peptide antigen via MHC class I
ERK1 and ERK2 cascade
Neuron apoptotic process
Positive regulation of gene expression
Regulation of cell growth
mRNA splicing, via spliceosome
Proteolysis
Protein export from nucleus
Peptide catabolic process
Antigen processing and presentation of endogenous peptide antigen via MHC class I
Antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent
Membrane protein ectodomain proteolysis
Oxidation-reduction process
Phosphatidylcholine biosynthetic process
Regulation of alternative mRNA splicing, via spliceosome
Intracellular transport of virus
Endosomal transport
Integrin-mediated signaling pathway
Positive regulation of I-kappaB kinase/NF-kappaB signaling
Response to insulin
Response to calcium ion
g
up
down
DE
Immune system
Innate immune system
Interferon signaling
Antiviral mechanism by interferon-stimulated genes
Interferon alpha/beta signaling
Interleukin-2 signaling pathway
T cell receptor regulation of apoptosis
BDNF signaling pathway
Toll receptor cascades
Interferon-gamma signaling pathway
Interleukin-1 regulation of extracellular matrix
Interleukin-1 signaling pathway
Interleukin-4 regulation of apoptosis
O-linked glycosylation of mucins
Negative regulators of RIG-I/MDA5 signaling
RIG-I/MDA5-mediated induction of IFN-alpha/beta pathways
Type II interferon signaling (interferon-gamma)
Leptin influence on immune response
Activation of IRF3/IRF7 mediated by TBK1/IKK epsilon
SMAD2/3 nuclear pathway
RIG-I-like receptor signaling pathway
Adaptive immune system
Leukocyte transendothelial migration
Phagosome
Glycolysis and gluconeogenesis
Lipid and lipoprotein metabolism
Carbohydrate metabolism
Protein processing in the endoplasmic reticulum
Protein metabolism
Antigen processing: cross presentation
MHC class II antigen presentation
up
down
DE
Biological oxidations
Cell adhesion molecules (CAMs)
HIF-1 transcriptional activity in hypoxia
HIF-1 degradation in normoxia
Apoptosis
Interleukin-5 regulation of apoptosis
Glycosaminoglycan metabolism
Insulin biosynthesis and processing
Beta-3 integrin cell surface interactions
ER-associated degradation (ERAD) pathway
Signal transduction by L1
Calcium regulation in the cardiac cell
Hypoxia-inducible factor in the cardiovascular system
Activated NOTCH1 signaling in the nucleus
Alpha-V beta-3 integrin/OPN pathway
Alpha-6 beta-4 integrin signaling pathway
CXCR4 signaling pathway
Zinc transporters
Iron uptake and transport
Nrdp1 controls ErbB3 receptor recycling
Antigen-activated B-cell receptor generation of 2nd messengers
T cell receptor signaling pathway
T cell signal transduction
Ion channel transport
Endothelial NOS activation
Calcium/calmodulin-dependent protein kinase activation
Cardiac muscle contraction
Signaling by insulin receptor
Influenza viral RNA transcription and replication
Interleukin-6 regulation of target genes
Interleukin-10 anti-inflammatory signaling pathway
h
up
down
DE
Ribosome biogenesis in eukaryotes
NOD-like receptor signaling pathway
IL-17 signaling pathway
Toll-like receptor signaling pathway
RIG-I-like receptor signaling pathway
MAPK signaling pathway
Antigen processing and presentation
B cell receptor signaling pathway
Endocytosis
Fc gamma R-mediated phagocytosis
Natural killer cell mediated cytotoxicity
HIF-1 signaling pathway
Cell adhesion molecules (CAMs)
TGF-beta signaling pathway
Notch signaling pathway
RNA degradation
PI3K-Akt signaling pathway
ABC transporters
Ubiquitin mediated proteolysis
p53 signaling pathway
up
down
DE
Leukocyte transendothelial migration
Thyroid hormone signaling pathway
Neurotrophin signaling pathway
MicroRNAs in cancer
Phagosome
Lysosome
Adherens junction
Fatty acid degradation
Apoptosis
Thyroid hormone synthesis
Oxidative phosphorylation
Type I diabetes mellitus
Insulin signaling pathway
Hippo signaling pathway
Inflammatory regulation of TRP channels
T cell receptor signaling pathway
TNF signaling pathway
Cytokine-cytokine receptor interaction
Th17 cell differentiation
Complement and coagulation cascades
P-value for enrichment
0.0
0.05
1.0
Significant
Non-Significant

(See figure on next page.)

Fig. 4 Comparison of the gene expression patterns in different SARS-CoV-2 infection models. **a** Venn-diagram showing the observed deregulated genes (with their respective control) in the different cell types. Enrichment analysis and comparison between deregulated genes in different SARS-CoV-2 infection models using **b** GOBP module, **c** Bioplanet pathway module, **d** KEGG pathway module. Selected significant terms are represented in heatmaps. Color scale/schemes are similar to Fig. 3

patients' samples. We spotted that both the healthy and infected samples have expressed these genes except the *DPP4* gene (Fig. 3e).

Genes that are deregulated due to SARS-CoV-2 initial infection site at nasopharyngeal region are not elucidated much so far. Here, we identified 1,614 differentially expressed genes within our reported four SARS-CoV-2 infected nasopharyngeal samples; among these differentially expressed genes, 558 genes were upregulated, and 1056 genes were downregulated (Additional file 4). Then we sought to discover the biological functions/pathways these deregulated genes might be involved in. To achieve this, we performed functional enrichment analyses with the observed deregulated genes using different ontology and pathway modules.

Several GOBP terms related to antiviral immune responses such as viral process, defense response to virus, innate immune response, inflammatory response, negative regulation of viral transcription, and negative regulation of viral genome replication were observed enriched for the upregulated genes (Fig. 3f, Additional file 5: Figure S2). Surprisingly, several other important antiviral defense related functions such as- apoptosis, and antigen processing and presentation were found enriched for downregulated genes (Fig. 3f).

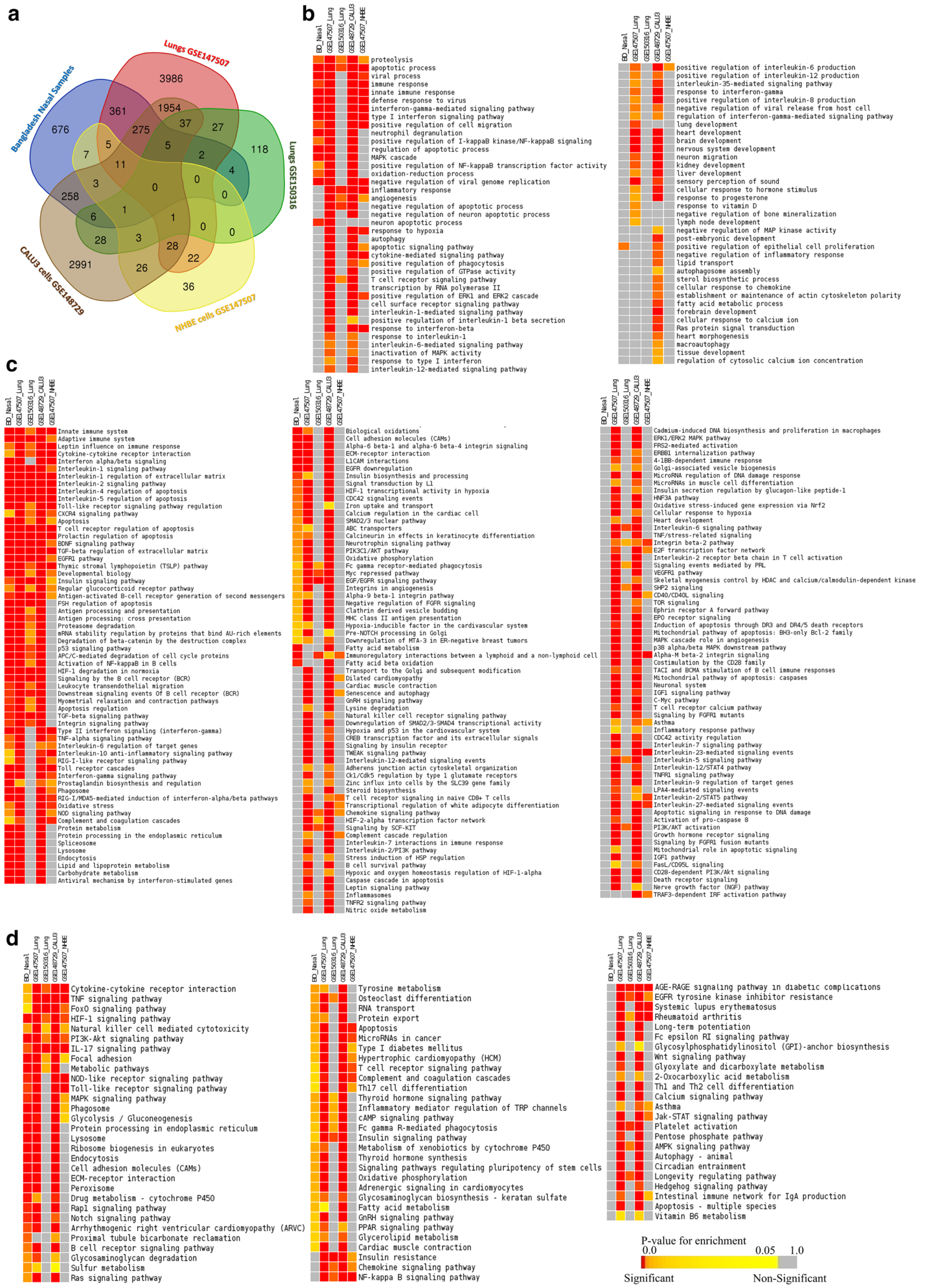
Similarly, this pattern was also observed for the functional enrichment using KEGG and Bioplanet pathways modules. Upregulated genes are observed involved in signaling pathways such as innate immune system, antiviral mechanism by interferon-stimulated genes, interleukin-2 signaling, interferon-gamma signaling, interferon alpha–beta signaling, antiviral mechanism by interferon stimulated genes, IL-17 signaling pathway, Toll-like receptor signaling pathway, RIG-I like receptor signaling pathway, and MAPK signaling pathway (Fig. 3g, h, Additional file 5: Figure S2). Strikingly, several important antiviral signaling pathways such as antigen processing and presentation, apoptosis, HIF-1 signaling pathway, Natural killer cell mediated cytotoxicity, phagosome, PI3K-Akt signaling pathway, Interleukin-6 regulation of target genes, and Interleukin-10 inflammatory signaling pathway were enriched for the downregulated genes (Fig. 3g, h). This unusual observation made us curious to search for a similar pattern of deregulated host responses in several other COVID-19 disease models.

Host responses observed in nasopharyngeal samples are significantly different compared to the other SARS-CoV-2 infections models

We sought to compare the host responses of our analyzed samples with several other different SARS-CoV-2 infection models (two different experiments containing lung biopsy samples from COVID-19 patients and two different SARS-CoV-2 infected cell lines). We performed functional enrichment analyses using differentially expressed genes from four other SARS-CoV-2 infection systems, and compared the enriched terms of our samples with these four other samples. Moreover, how the host responds differently in different tissue types were also evaluated. To achieve these goals, we identified the differentially expressed genes across these different samples and systematically compared the enrichment results of those deregulated genes.

Using the similar parameterization of the differential gene expression analyses, we identified 6714 genes in lung cells (GSE147507), 232 genes in lung cells (GSE150316), 143 genes in NHBE cells (GSE147507), and 5637 genes in Calu-3 cells (GSE148729) as differentially expressed compared to their respective healthy controls (Additional file 6). Significant proportions of the deregulated genes detected in our nasopharyngeal samples are also found deregulated in lung (GSE147507) and Calu-3 cells (GSE148729) samples (Fig. 4a), while a small number of our samples' deregulated genes were also observed deregulated in rest of the two samples used (Fig. 4a).

Enrichment analysis using these deregulated genes suggested the host response differences among the different infection systems used (Fig. 4b–d). Upon the analysis, only a few GOBP terms were found enriched for both our samples, lung (GSE147507), and Calu-3 cells (GSE148729) samples, such as viral process, immune response, innate immune response, defense response to virus, and interferon signaling (Fig. 4b). However, genes in many important antiviral immune response related functions were not significantly enriched in nasopharyngeal samples but were enriched for the lung (GSE147507), and Calu-3 cells (GSE148729) samples; these processes are autophagy, apoptotic signaling pathway, interleukin-6 mediated signaling pathway, interleukin-12 mediated signaling pathway, cytokine-mediated signaling pathway, and inflammatory


a
Lungs GSE147507
Bangladesh Nasal Samples
Lungs GSE150316
CALU3 cells GSE148729
NHBE cells GSE147507
3986
361
1954
676
275
37
27
5
5
7
11
2
118
4
3
0
0
258
0
1
6
1
0
0
28
3
28
2991
26
22
36
b
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
proteolysis
apoptotic process
viral process
immune response
innate immune response
defense response to virus
interferon-gamma-mediated signaling pathway
type I interferon signaling pathway
positive regulation of cell migration
neutrophil degranulation
positive regulation of I-kappaB kinase/NF-kappaB signaling
regulation of apoptotic process
MAPK cascade
positive regulation of NF-kappaB transcription factor activity
oxidation-reduction process
negative regulation of viral genome replication
inflammatory response
angiogenesis
negative regulation of apoptotic process
negative regulation of neuron apoptotic process
neuron apoptotic process
response to hypoxia
autophagy
apoptotic signaling pathway
cytokine-mediated signaling pathway
positive regulation of phagocytosis
positive regulation of GTPase activity
T cell receptor signaling pathway
transcription by RNA polymerase II
positive regulation of ERK1 and ERK2 cascade
cell surface receptor signaling pathway
interleukin-1-mediated signaling pathway
positive regulation of interleukin-1 beta secretion
response to interferon-beta
response to interleukin-1
interleukin-6-mediated signaling pathway
inactivation of MAPK activity
response to type I interferon
interleukin-12-mediated signaling pathway
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
positive regulation of interleukin-6 production
positive regulation of interleukin-12 production
interleukin-35-mediated signaling pathway
response to interferon-gamma
positive regulation of interleukin-8 production
negative regulation of viral release from host cell
regulation of interferon-gamma-mediated signaling pathway
lung development
heart development
brain development
nervous system development
neuron migration
kidney development
liver development
sensory perception of sound
cellular response to hormone stimulus
response to progesterone
response to vitamin D
negative regulation of bone mineralization
lymph node development
negative regulation of MAP kinase activity
post-embryonic development
positive regulation of epithelial cell proliferation
negative regulation of inflammatory response
lipid transport
autophagosome assembly
sterol biosynthetic process
cellular response to chemokine
establishment or maintenance of actin cytoskeleton polarity
fatty acid metabolic process
forebrain development
cellular response to calcium ion
Ras protein signal transduction
heart morphogenesis
macroautophagy
tissue development
regulation of cytosolic calcium ion concentration
c
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
Innate immune system
Adaptive immune system
Leptin influence on immune response
Cytokine-cytokine receptor interaction
Interferon alpha/beta signaling
Interleukin-1 signaling pathway
Interleukin-1 regulation of extracellular matrix
Interleukin-2 signaling pathway
Interleukin-4 regulation of apoptosis
Interleukin-5 regulation of apoptosis
Toll-like receptor signaling pathway regulation
CXCR4 signaling pathway
Apoptosis
T cell receptor regulation of apoptosis
Prolactin regulation of apoptosis
BDNF signaling pathway
TGF-beta regulation of extracellular matrix
EGFR1 pathway
Thymic stromal lymphopoietin (TSLP) pathway
Developmental biology
Insulin signaling pathway
Regular glucocorticoid receptor pathway
Antigen-activated B-cell receptor generation of second messengers
FSH regulation of apoptosis
Antigen processing and presentation
Antigen processing: cross presentation
Proteasome degradation
mRNA stability regulation by proteins that bind AU-rich elements
Degradation of beta-catenin by the destruction complex
p53 signaling pathway
APC/C-mediated degradation of cell cycle proteins
Activation of NF-kappaB in B cells
HIF-1 degradation in normoxia
Signaling by the B cell receptor (BCR)
Leukocyte transendothelial migration
Downstream signaling events Of B cell receptor (BCR)
Myometrial relaxation and contraction pathways
Apoptosis regulation
TGF-beta signaling pathway
Integrin signaling pathway
Type II interferon signaling (interferon-gamma)
TNF-alpha signaling pathway
Interleukin-6 regulation of target genes
Interleukin-10 anti-inflammatory signaling pathway
RIG-I-like receptor signaling pathway
Toll receptor cascades
Interferon-gamma signaling pathway
Prostaglandin biosynthesis and regulation
Phagosome
RIG-I/MDA5-mediated induction of interferon-alpha/beta pathways
Oxidative stress
NOD signaling pathway
Complement and coagulation cascades
Protein metabolism
Protein processing in the endoplasmic reticulum
Spliceosome
Lysosome
Endocytosis
Lipid and lipoprotein metabolism
Carbohydrate metabolism
Antiviral mechanism by interferon-stimulated genes
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
Biological oxidations
Cell adhesion molecules (CAMs)
Alpha-6 beta-1 and alpha-6 beta-4 integrin signaling
ECM-receptor interaction
L1CAM interactions
EGFR downregulation
Insulin biosynthesis and processing
Signal transduction by L1
HIF-1 transcriptional activity in hypoxia
CDC42 signaling events
Iron uptake and transport
Calcium regulation in the cardiac cell
SMAD2/3 nuclear pathway
ABC transporters
Calcineurin in effects in keratinocyte differentiation
Neurotrophin signaling pathway
PIK3C1/AKT pathway
Oxidative phosphorylation
Fc gamma receptor-mediated phagocytosis
Myc repressed pathway
EGF/EGFR signaling pathway
Integrins in angiogenesis
Alpha-9 beta-1 integrin pathway
Negative regulation of FGFR signaling
Clathrin derived vesicle budding
MHC class II antigen presentation
Hypoxia-inducible factor in the cardivascular system
Pre-NOTCH processing in Golgi
Downregulation of MTA-3 in ER-negative breast tumors
Fatty acid metabolism
Immunoregulatory interactions between a lymphoid and a non-lymphoid cell
Fatty acid beta oxidation
Transport to the Golgi and subsequent modification
Dilated cardiomyopathy
Cardiac muscle contraction
Senescence and autophagy
GnRH signaling pathway
Lysine degradation
Natural killer cell receptor signaling pathway
Downregulation of SMAD2/3-SMAD4 transcriptional activity
Hypoxia and p53 in the cardiovascular system
CREB transcription factor and its extracellular signals
Signaling by insulin receptor
TWEAK signaling pathway
Interleukin-12-mediated signaling events
Adherens junction actin cytoskeletal organization
Ck1/Cdk5 regulation by type 1 glutamate receptors
Zinc influx into cells by the SLC39 gene family
Steroid biosynthesis
T cell receptor signaling in naive CD8+ T cells
Transcriptional regulation of white adipocyte differentiation
Chemokine signaling pathway
HIF-2-alpha transcription factor network
Signaling by SCF-KIT
Complement cascade regulation
Interleukin-7 interactions in immune response
Interleukin-2/PI3K pathway
Stress induction of HSP regulation
B cell survival pathway
Hypoxic and oxygen homeostasis regulation of HIF-1-alpha
Caspase cascade in apoptosis
Leptin signaling pathway
Inflammasomes
TNFR2 signaling pathway
Nitric oxide metabolism
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
Cadmium-induced DNA biosynthesis and proliferation in macrophages
ERK1/ERK2 MAPK pathway
FRS2-mediated activation
ERBB1 internalization pathway
4-1BB-dependent immune response
Golgi-associated vesicle biogenesis
MicroRNA regulation of DNA damage response
MicroRNAs in muscle cell differentiation
Insulin secretion regulation by glucagon-like peptide-1
HNF3A pathway
Oxidative stress-induced gene expression via Nrf2
Cellular response to hypoxia
Heart development
Interleukin-6 signaling pathway
TNF/stress-related signaling
Integrin beta-2 pathway
E2F transcription factor network
Interleukin-2 receptor beta chain in T cell activation
Signaling events mediated by PRL
VEGFR1 pathway
Skeletal myogenesis control by HDAC and calcium/calmodulin-dependent kinase
SHP2 signaling
CD40/CD40L signaling
TOR signaling
Ephrin receptor A forward pathway
EPO receptor signaling
Induction of apoptosis through DR3 and DR4/5 death receptors
Mitochondrial pathway of apoptosis: BH3-only Bcl-2 family
MAPK cascade role in angiogenesis
p38 alpha/beta MAPK downstream pathway
Alpha-M beta-2 integrin signaling
Costimulation by the CD28 family
TACI and BCMA stimulation of B cell immune responses
Mitochondrial pathway of apoptosis: caspases
Neuronal system
IGF1 signaling pathway
C-Myc pathway
T cell receptor calcium pathway
Signaling by FGFR1 mutants
Asthma
Inflammatory response pathway
CDC42 activity regulation
Interleukin-7 signaling pathway
Interleukin-23-mediated signaling events
Interleukin-5 signaling pathway
Interleukin-12/STAT4 pathway
TNFR1 signaling pathway
Interleukin-9 regulation of target genes
LPA4-mediated signaling events
Interleukin-2/STAT5 pathway
Interleukin-27-mediated signaling events
Apoptotic signaling in response to DNA damage
Activation of pro-caspase 8
PI3K/AKT activation
Growth hormone receptor signaling
Signaling by FGFR1 fusion mutants
Mitochondrial role in apoptotic signaling
IGF1 pathway
FasL/CD95L signaling
CD28-dependent PI3K/Akt signaling
Death receptor signaling
Nerve growth factor (NGF) pathway
TRAF3-dependent IRF activation pathway
d
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
Cytokine-cytokine receptor interaction
TNF signaling pathway
FoxO signaling pathway
HIF-1 signaling pathway
Natural killer cell mediated cytotoxicity
PI3K-Akt signaling pathway
IL-17 signaling pathway
Focal adhesion
Metabolic pathways
NOD-like receptor signaling pathway
Toll-like receptor signaling pathway
MAPK signaling pathway
Phagosome
Glycolysis / Gluconeogenesis
Protein processing in endoplasmic reticulum
Lysosome
Ribosome biogenesis in eukaryotes
Endocytosis
Cell adhesion molecules (CAMs)
ECM-receptor interaction
Peroxisome
Drug metabolism - cytochrome P450
Rap1 signaling pathway
Notch signaling pathway
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Proximal tubule bicarbonate reclamation
B cell receptor signaling pathway
Glycosaminoglycan degradation
Sulfur metabolism
Ras signaling pathway
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
Tyrosine metabolism
Osteoclast differentiation
RNA transport
Protein export
Apoptosis
MicroRNAs in cancer
Type I diabetes mellitus
Hypertrophic cardiomyopathy (HCM)
T cell receptor signaling pathway
Complement and coagulation cascades
Th17 cell differentiation
Thyroid hormone signaling pathway
Inflammatory mediator regulation of TRP channels
cAMP signaling pathway
Fc gamma R-mediated phagocytosis
Insulin signaling pathway
Metabolism of xenobiotics by cytochrome P450
Thyroid hormone synthesis
Signaling pathways regulating pluripotency of stem cells
Oxidative phosphorylation
Adrenergic signaling in cardiomyocytes
Glycosaminoglycan biosynthesis - keratan sulfate
Fatty acid metabolism
GnRH signaling pathway
PPAR signaling pathway
Glycerolipid metabolism
Cardiac muscle contraction
Insulin resistance
Chemokine signaling pathway
NF-kappa B signaling pathway
BD_Nasal
GSE147507_Lung
GSE150316_Lung
GSE148729_CALU3
GSE147507_NHBE
AGE-RAGE signaling pathway in diabetic complications
EGFR tyrosine kinase inhibitor resistance
Systemic lupus erythematosus
Rheumatoid arthritis
Long-term potentiation
Fc epsilon RI signaling pathway
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis
Wnt signaling pathway
Glyoxylate and dicarboxylate metabolism
2-Oxocarboxylic acid metabolism
Th1 and Th2 cell differentiation
Calcium signaling pathway
Asthma
Jak-STAT signaling pathway
Platelet activation
Pentose phosphate pathway
AMPK signaling pathway
Autophagy - animal
Circadian entrainment
Longevity regulating pathway
Hedgehog signaling pathway
Intestinal immune network for IgA production
Apoptosis - multiple species
Vitamin B6 metabolism
P-value for enrichment
0.0
0.05
1.0
Significant
Non-Significant

(See figure on next page.)

Fig. 5 Gene expression analysis using different SARS-CoV-2 infection models. **a** Variance plot, **b** Sample to sample distance plot, **c** Principal component analysis plot, **d** Clustered heatmap of the count matrix of the normalized RNA-seq reads of different SARS-CoV-2 infection samples using to 50 genes. **e** Gene expression heatmap showing global gene expression profiles in the individual infected samples of the various infection system. Heatmap is clustered based on Pearson's distance with genes that vary across the sample, leaving out genes that do not vary significantly

response (Fig. 4b). Moreover, processes such as response to hypoxia, response to vitamin-D, and lung development were also not enriched for the deregulated genes in nasal samples (Fig. 4b).

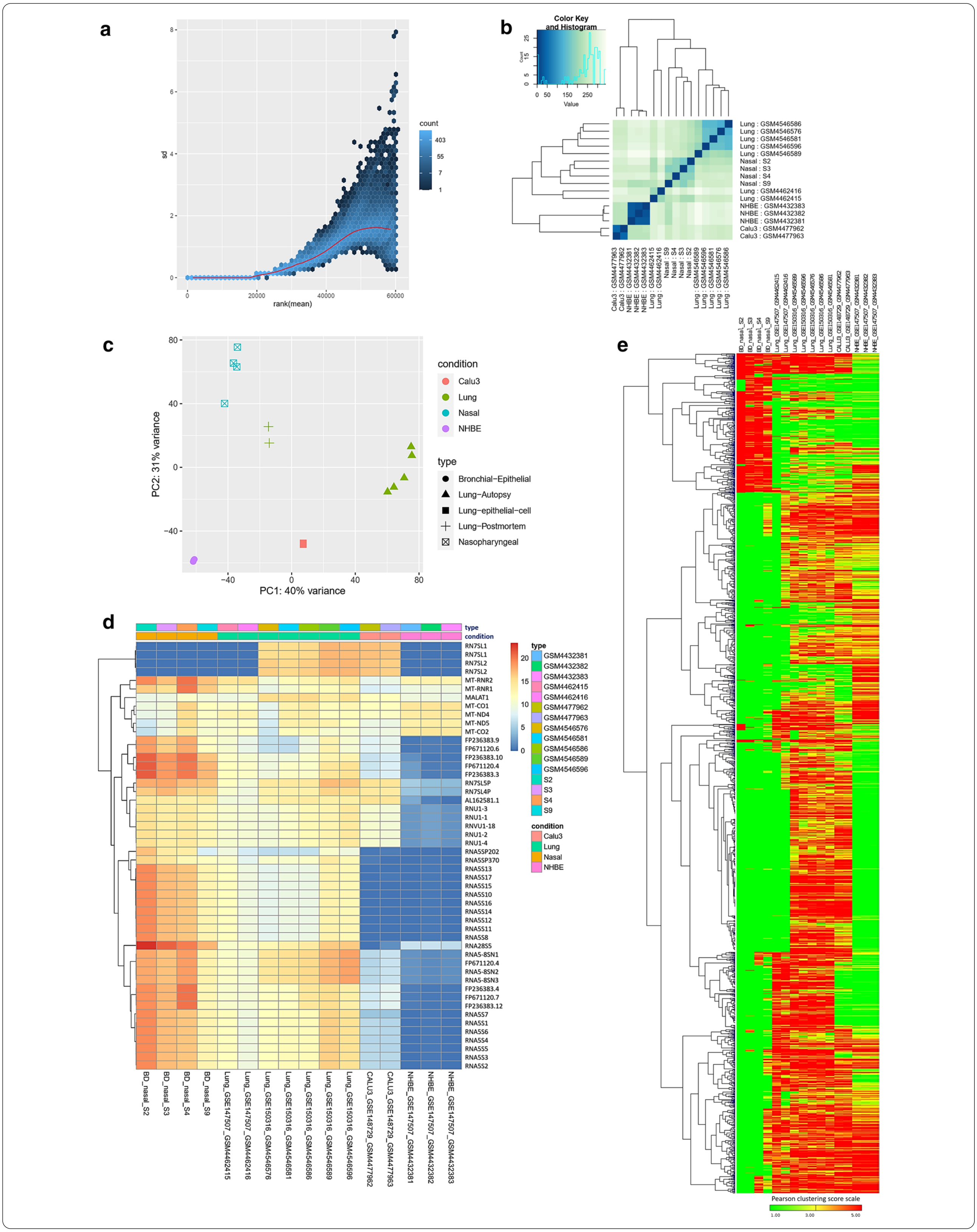
We noticed several commonly enriched important immune signaling pathways for most of the samples used for the comparison (Fig. 4c, d), such as adaptive immune system, innate immune system, interferon signaling, apoptosis, Toll-like receptor signaling pathway regulation, antigen processing and presentation, integrin signaling pathway, RIG-I like receptor signaling pathway, and phagosomes (Fig. 4c, d, Additional file 7: Figure S3); however, pathways such as JAK-STAT signaling pathway, Natural killer cell mediated cytotoxicity, NF-κB signaling pathway, asthma, PI3K-Akt pathway, cellular response to hypoxia, inflammasomes, and inflammatory response pathway (Fig. 4c, d, Additional file 7: Figure S3) were not enriched for the deregulated genes of our nasopharyngeal samples. These results suggest that host responses observed in nasopharyngeal samples have a different host response compared to the other infection systems. However, the diffenences observed in the infected cell lines' transcriptomes might be the resultant effects of the inherent variability of these cells compared to the nasal epithelial cells or the lung cells. Therefore, to unveil the mystery behind this observation, we further analyzed these data to compare the COVID-19 patients' nasal and lung gene expression patterns for different specific functionalities.

Significant gene expression differences were spotted between the nasopharyngeal samples and lung biopsy samples

We compared the normalized read counts of each infected nasal and lung samples without integrating the respective controls to shed insights on the differences in gene expression patterns between the individual samples and tissues. A constant standard deviation was observed for the normalized read counts of the infected samples (Fig. 5a) indicating the acceptability of the normalized reads for analysis. From the sample to sample distance clustering, principal component analysis, and clustered heatmap of the count matrix with top 50 genes, we observed that gene expression profiles of our nasopharyngeal samples are more relevant to that of lung samples; whereas, high level of variance was observed between the gene expression counts of the cell lines and primary nasopharyngeal samples (Fig. 5b–d). PCA analysis (Fig. 5c) also suggests that cell line data are quite different than primary samples data. Furthermore, we had a similar observation from the clustered normalized read counts of the samples based on Pearson's correlation distance with all genes that vary across samples (Fig. 5e). We then narrowed down our searches to the sample level gene expression profiles of several COVID-19 related important biological functions within these samples (Fig. 6), to understand the gene expression similarities and dissimilarities among these infections systems, specially comparing nasal and lung tissues.

Genes related to integrins and integrin signaling pathway are highly expressed in lung samples compared to the nasopharyngeal samples

Though some previous reports [37, 38] suggested an important aspect of integrins in SARS-CoV-2 pathogenesis, precise information on which particular integrins are deregulated and how virus interactions might modulate them remained unclear. Therefore, we sought to find out the expression profiles of integrin related genes in different COVID-19 infection models at sample level. RGD motif of the spike protein of SARS-CoV-2 can bind the integrins and this motif is placed near to the ACE2-receptor binding motif [37]. Moreover, evidence of integrin domain binding was also reported for SARS-CoV [39]. Therefore, we sought to discover the expression profiles of the integrin related genes in different SARS-CoV-2 infection models. To accomplish this, we filtered out the integrin and integrin signaling related genes (Additional file 8) within the terms of the GOBP, KEGG pathway, and Bioplanet pathway modules that we used for enrichment analysis. Intriguingly, we observed that the genes related to integrins and integrin signaling such as *ITGAV, ITGA6, ITGB7, ITGB3, ITGA2B, ITGA5, ITGA6, ITGA9, ITGA4, ITGAE,* and *ITGA8* were highly expressed in analyzed lung samples, and the lowest number of these genes were expressed in the nasopharyngeal samples (Fig. 6a, b, Additional file 9: Figure S4A). Based on these observations, we can assume that overexpression of integrins and integrin signaling related genes in the lungs might provide the virus a competitive edge in invading the lung cells more efficiently compared to the cells of the nasopharynx and respiratory tracts.


a
sd
rank(mean)
0
20000
40000
60000
0
2
4
6
8
count
403
55
7
1
b
Color Key and Histogram
Count
Value
0 50 150 250
0 5 15 25
Lung : GSM4546586
Lung : GSM4546576
Lung : GSM4546581
Lung : GSM4546596
Lung : GSM4546589
Nasal : S2
Nasal : S3
Nasal : S4
Nasal : S9
Lung : GSM4462416
Lung : GSM4462415
NHBE : GSM4432383
NHBE : GSM4432382
NHBE : GSM4432381
Calu3 : GSM4477962
Calu3 : GSM4477963
c
PC1: 40% variance
PC2: 31% variance
-40
0
40
80
condition
Calu3
Lung
Nasal
NHBE
type
Bronchial−Epithelial
Lung−Autopsy
Lung−epithelial−cell
Lung−Postmortem
Nasopharyngeal
d
type
condition
RN7SL1
RN7SL1
RN7SL2
RN7SL2
MT-RNR2
MT-RNR1
MALAT1
MT-CO1
MT-ND4
MT-ND5
MT-CO2
FP236383.9
FP671120.6
FP236383.10
FP671120.4
FP236383.3
RN7SL5P
RN7SL4P
AL162581.1
RNU1-3
RNU1-1
RNVU1-18
RNU1-2
RNU1-4
RNA5SP202
RNA5SP370
RNA5S13
RNA5S17
RNA5S15
RNA5S10
RNA5S16
RNA5S14
RNA5S12
RNA5S11
RNA5S8
RNA28S5
RNA5-8SN1
FP671120.4
RNA5-8SN2
RNA5-8SN3
FP236383.4
FP671120.7
FP236383.12
RNA5S7
RNA5S1
RNA5S6
RNA5S4
RNA5S5
RNA5S3
RNA5S2
20
15
10
5
0
GSM4432381
GSM4432382
GSM4432383
GSM4462415
GSM4462416
GSM4477962
GSM4477963
GSM4546576
GSM4546581
GSM4546586
GSM4546589
GSM4546596
S2
S3
S4
S9
Calu3
Lung
Nasal
NHBE
BD_nasal_S2
BD_nasal_S3
BD_nasal_S4
BD_nasal_S9
Lung_GSE147507_GSM4462415
Lung_GSE147507_GSM4462416
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
CALU3_GSE148729_GSM4477962
CALU3_GSE148729_GSM4477963
NHBE_GSE147507_GSM4432381
NHBE_GSE147507_GSM4432382
NHBE_GSE147507_GSM4432383
e
Pearson clustering score scale
1.00
3.00
5.00
]

(See figure on next page.)

Fig. 6 Heatmaps representing the sample level absolute expression of Integrin related genes. **a** across the different SARS-CoV-2 infection models, **b** in only nasopharyngeal samples and lung samples; Cytokine signaling related genes **c** across the different SARS-CoV-2 infection models, **d** in only nasopharyngeal samples and lung samples; and Inflammation related genes **e** across the different SARS-CoV-2 infection models, **f** in only nasopharyngeal samples and lung samples; **g** Expression profiles of genes encoding SARS-CoV-2 receptor and entry associated proteins. Normalized (DESeq2) Log_2 converted read counts are considered as the expression values of the genes and represented in a color-coded scale; Color towards red indicating higher expression, while color towards green indicating little to no expression

Cytokine and inflammatory signaling genes are overexpressed in lung samples

Aberrant cytokine stimulation and inflammatory responses are thought to be the major contributor to pathogenic lung damages in severely affected COVID-19 patients [40, 41]. We wanted to find out whether the genes related to cytokine signaling and inflammation have differential expression profiles in lung cells compared to the other infection systems. We extracted and compared the gene expression values of the genes related to these two terms (Additional file 8). We are not surprised to observe that the genes of these two major contributing events of COVID-19 lung pathobiology are significantly overexpressed in lung samples compared to the rest of the SARS-CoV-2 infected cell types (Fig. 6c–f, Additional file 9: Figure S4B-C). Particularly, the analyzed nasopharyngeal samples have very low expression values for the cytokine and inflammatory signaling genes such as *CCL4, TNFA, IL6, IL1A, CCL2, CXCL2, IFN,* and *CCR1* (Fig. 6c–f). Therefore, these observations are fueling the preexisting supposition of the roles of enhanced cytokine, and inflammatory signaling for worsening the disease condition in patients with SARS-CoV-2 infected lungs.

A differential gene expression profile was detected for the SARS-CoV-2 entry receptors/associated proteins in different infection models

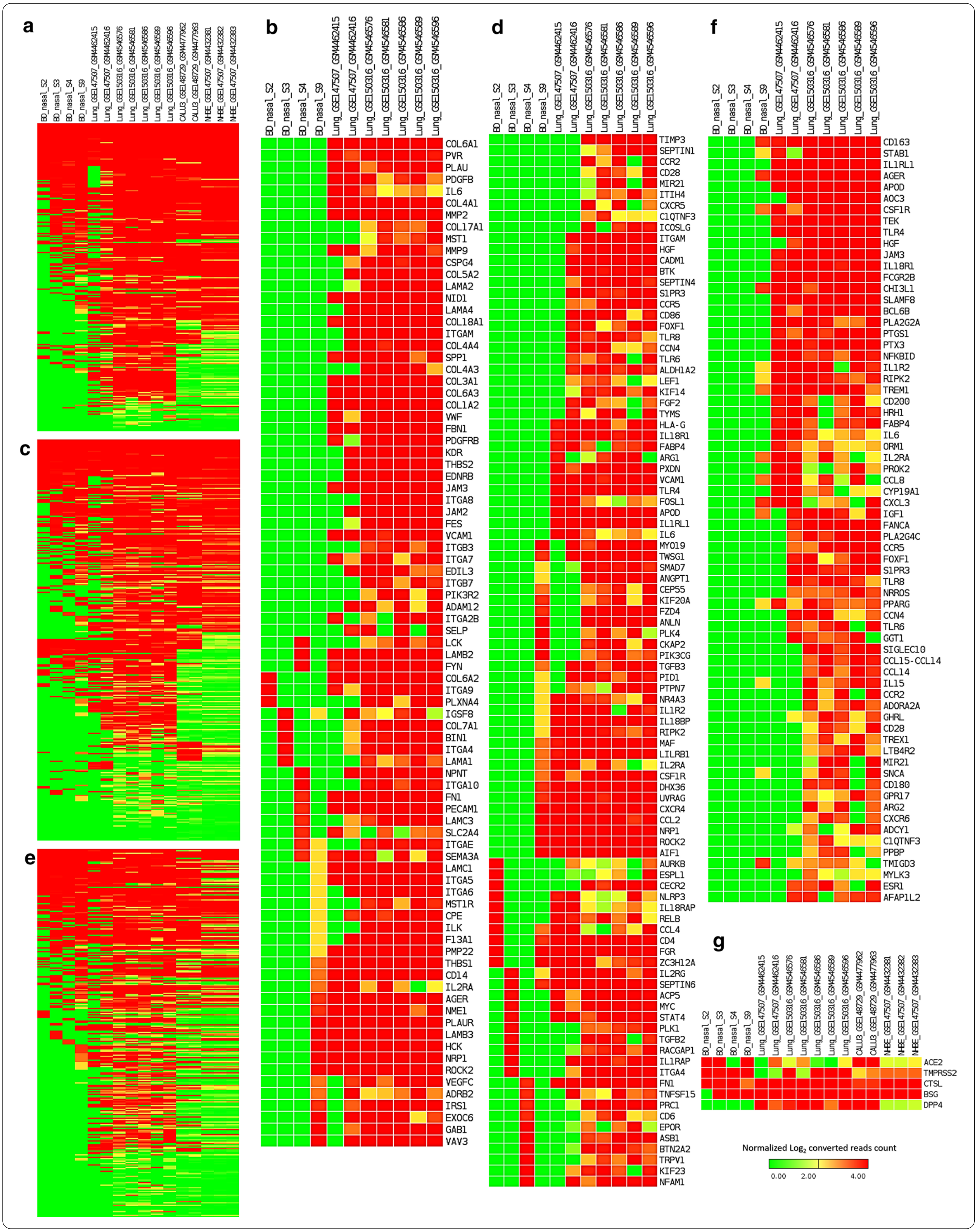
Expression of receptor protein ACE2 and entry associated proteins such as TMPRSS2, BSG, CTSL, DPP4 on the cell surface of the host is essential for the invasion of SARS-CoV-2 [36]. Moreover, ACE2 overexpression is thought to increase the infection potentiality of SARS-CoV-2 [42]. Furthermore, Kuba et al. demonstrated the potential role of ACE2 in SARS-CoV induced lung injury [43]. So, we ventured to check the gene expression levels of ACE2 and the other entry associated proteins in the different SARS-CoV-2 infected cells. Surprisingly, we observed that the *ACE2* gene was not expressed in high levels in lung samples as speculated (Fig. 6g). However, gene expression levels of the other entry associated proteins were higher in lung samples (Fig. 6g). Nonetheless, in few of the lung samples, the *TMPRSS2* gene was not expressed in higher amounts (Fig. 6g). Interestingly, we have not detected any expression of *DPP4* gene within the reported nasopharyngeal samples (Fig. 6g).

Inflammatory immune responses were several folds higher in lungs than the nasopharynx of COVID-19 patients

From our previous observations, it was evident that COVID-19 patient's lung responds to the viral infection differently compared to the epithelial cells of nasopharynx. We then sought to figure out the specific genes and biological functions/signaling pathways which have this differential pattern. We achieved this by designing a multifactorial differential gene expression analysis using a generalized linear model (GLM) [44]; in which we compared the fold changes of every differentially expressed gene in our nasopharyngeal samples and lung (GSE150316) samples, to discover how many folds lung is alternatively expressing the genes than nasopharynx in COVID-19.

Firstly, we analyzed the suitability of the data for this design and observed no irregularities between the data used (Fig. 7a–d). Moreover, upon this multifactorial differential gene expression analysis, we observed an acceptable common biological coefficient of variation; this variation decreases significantly as the expression values increases (Fig. 7e). From the MA plot, we observed a very high amount of the significantly ($p\text{-value} < 0.05$) several fold upregulated and downregulated genes in lungs compared to nasopharyngeal samples (Fig. 7f). We detected 807 upregulated and 298 downregulated genes in lungs compared to the nasopharyngeal samples (Additional file 10). Interestingly, we noticed the highly upregulated integrin and integrin signaling genes in lungs compared to the nasal samples (Fig. 7g) which are consistent with our previous observations. Modulatory roles of integrins are well established in acute lung damages [45]. Similarly, aberrant expression of genes involved in integrin signaling can also provoke acute lung injuries, namely-*ADAM15* [46], *SDC1* [47], *CD14* [48], *CD47* [49], *CD9* [50], *HMGB1* [51], *ITA6* [52], and *ITAV* [53]. Therefore, SARS-CoV-2 infection induced deregulation of these genes might be contributing towards the worsening of the normal pathobiology and functionality of lungs in COVID-19.

We then performed functional enrichment analysis to hunt down the signaling pathways which are differentially


a
BD_nasal_S2
BD_nasal_S3
BD_nasal_S4
BD_nasal_S9
Lung_GSE147507_GSM4462415
Lung_GSE147507_GSM4462416
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
CALU3_GSE148729_GSM4477962
CALU3_GSE148729_GSM4477963
NHBE_GSE147507_GSM4432381
NHBE_GSE147507_GSM4432382
NHBE_GSE147507_GSM4432383
c
e
b
BD_nasal_S2
BD_nasal_S3
BD_nasal_S4
BD_nasal_S9
Lung_GSE147507_GSM4462415
Lung_GSE147507_GSM4462416
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
COL6A1
PVR
PLAU
PDGFB
IL6
COL4A1
MMP2
COL17A1
MST1
MMP9
CSPG4
COL5A2
LAMA2
NID1
LAMA4
COL18A1
ITGAM
COL4A4
SPP1
COL4A3
COL3A1
COL6A3
COL1A2
VWF
FBN1
PDGFRB
KDR
THBS2
EDNRB
JAM3
ITGA8
JAM2
FES
VCAM1
ITGB3
ITGA7
EDIL3
ITGB7
PIK3R2
ADAM12
ITGA2B
SELP
LCK
LAMB2
FYN
COL6A2
ITGA9
PLXNA4
IGSF8
COL7A1
BIN1
ITGA4
LAMA1
NPNT
ITGA10
FN1
PECAM1
LAMC3
SLC2A4
ITGAE
SEMA3A
LAMC1
ITGA5
ITGA6
MST1R
CPE
ILK
F13A1
PMP22
THBS1
CD14
IL2RA
AGER
NME1
PLAUR
LAMB3
HCK
NRP1
ROCK2
VEGFC
ADRB2
IRS1
EXOC6
GAB1
VAV3
d
BD_nasal_S2
BD_nasal_S3
BD_nasal_S4
BD_nasal_S9
Lung_GSE147507_GSM4462415
Lung_GSE147507_GSM4462416
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
TIMP3
SEPTIN1
CCR2
CD28
MIR21
ITIH4
CXCR5
C1QTNF3
ICOSLG
ITGAM
HGF
CADM1
BTK
SEPTIN4
S1PR3
CCR5
CD86
FOXF1
TLR8
CCN4
TLR6
ALDH1A2
LEF1
KIF14
FGF2
TYMS
HLA-G
IL18R1
FABP4
ARG1
PXDN
VCAM1
TLR4
FOSL1
APOD
IL1RL1
IL6
MYO19
TWSG1
SMAD7
ANGPT1
CEP55
KIF20A
FZD4
ANLN
PLK4
CKAP2
PIK3CG
TGFB3
PID1
PTPN7
NR4A3
IL1R2
IL18BP
RIPK2
MAF
LILRB1
IL2RA
CSF1R
DHX36
UVRAG
CXCR4
CCL2
NRP1
ROCK2
AIF1
AURKB
ESPL1
CECR2
NLRP3
IL18RAP
RELB
CCL4
CD4
FGR
ZC3H12A
IL2RG
SEPTIN6
ACP5
MYC
STAT4
PLK1
TGFB2
RACGAP1
IL1RAP
ITGA4
FN1
TNFSF15
PRC1
CD6
EPOR
ASB1
BTN2A2
TRPV1
KIF23
NFAM1
f
BD_nasal_S2
BD_nasal_S3
BD_nasal_S4
BD_nasal_S9
Lung_GSE147507_GSM4462415
Lung_GSE147507_GSM4462416
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
CD163
STAB1
IL1RL1
AGER
APOD
AOC3
CSF1R
TEK
TLR4
HGF
JAM3
IL18R1
FCGR2B
CHI3L1
SLAMF8
BCL6B
PLA2G2A
PTGS1
PTX3
NFKBID
IL1R2
RIPK2
TREM1
CD200
HRH1
FABP4
IL6
ORM1
IL2RA
PROK2
CCL8
CYP19A1
CXCL3
IGF1
FANCA
PLA2G4C
CCR5
FOXF1
S1PR3
TLR8
NRROS
PPARG
CCN4
TLR6
GGT1
SIGLEC10
CCL15-CCL14
CCL14
IL15
CCR2
ADORA2A
GHRL
CD28
TREX1
LTB4R2
MIR21
SNCA
CD180
GPR17
ARG2
CXCR6
ADCY1
C1QTNF3
PPBP
TMIGD3
MYLK3
ESR1
AFAP1L2
g
BD_nasal_S2
BD_nasal_S3
BD_nasal_S4
BD_nasal_S9
Lung_GSE147507_GSM4462415
Lung_GSE147507_GSM4462416
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
CALU3_GSE148729_GSM4477962
CALU3_GSE148729_GSM4477963
NHBE_GSE147507_GSM4432381
NHBE_GSE147507_GSM4432382
NHBE_GSE147507_GSM4432383
ACE2
TMPRSS2
CTSL
BSG
DPP4
Normalized Log$_2$ converted reads count
0.00
2.00
4.00

(See figure on next page.)

Fig. 7 Multifactorial differential gene expression analysis using differentially expressed COVID-19 lung and nasal data. **a** Variance plot, **b** Sample to sample distance plot, **c** Principal component analysis plot, **d** Clustered heatmap of the count matrix of the normalized RNA-seq reads (top 50 genes) of the SARS-CoV-2 infected nasopharyngeal and lung samples. **e** Common dispersion plot or the biological coefficient of variation plot. Here we are estimating the dispersion. The square root of the common dispersion gives the coefficient of variation of biological variation. Here the coefficient of biological variation is around 0.8. **f** MA plot. Plot log-fold change against log-counts per million, with DE genes are highlighted. The blue lines indicate twofold changes. Red and blue points indicate genes with P-value less than 0.05. **g** Expression profiles of genes encoding Integrins. Log_2 (fold change) values are considered as the expression values of the genes and represented in a color-coded scale; Color towards red indicating higher expression, while color towards green indicating little to no expression

expressed in lungs compared to the nasopharyngeal cells. These enrichment analyses revealed that biological functions such as viral process, and antigen processing and presentation were highly upregulated, function such as regulation of gene silencing by miRNA was found downregulated in lungs compared to the nasopharyngeal cells (Fig. 8a). Furthermore, pathways that provide antiviral immunity such as apoptosis, phagosome, antigen processing and presentation, adaptive immune system, innate immune system, interferon signaling, different interleukin signaling, and cytokine signaling in immune system were highly upregulated in lungs compared to the nasopharyngeal samples (Fig. 8b–d). Despite having the antiviral protective roles, hyperactivity from these pathways can significantly worsen the COVID-19 patient's overall lung functionality which can be further complicated with progressive and permanent lung damage.

Previously, it was reported that transcription factors can contribute to many inflammatory lung diseases [54, 55] which have similar lung characteristics observed in COVID-19. In this context, we identified the highly expressed transcription factors in lungs by comparing their respective expression values in nasopharyngeal samples (Fig. 8e). Among these, transcription factors such as CBP [54], CEBP [56], NFAT [54], ATF3 [57], GATA6 [58], HDAC2 [59], and TCF12 [60] have significant roles in lung's overall functionality, acute lung injury and antiviral response mechanism in lungs.

SARS-CoV-2 integrates its proteins in regulating the host antiviral immune responses

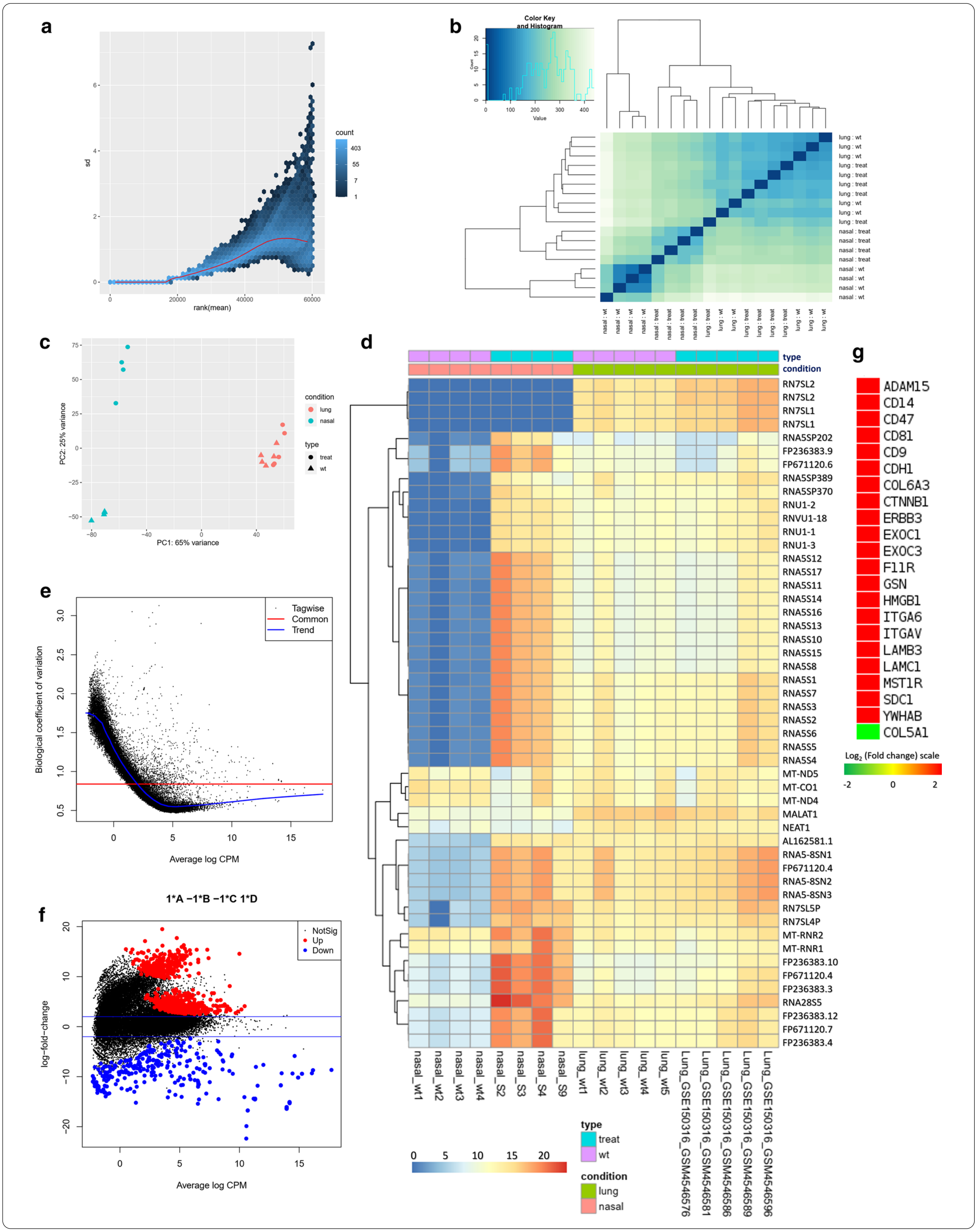
As we have observed the differential host responses in COVID-19 nasopharyngeal samples, then we sought to interconnect the virus-host interplay in those host responses. We first analyzed how many of the virus interacting host proteins' genes reported by Gordon et al. [21] are differentially expressed in our reported nasopharyngeal samples. Only 51 genes of those proteins are found deregulated in our nasopharyngeal samples (Fig. 9a). We then constructed a network interlinking the virus-host protein–protein interaction data from Gordon et al. [21] along with the deregulated genes from the nasopharyngeal samples (Fig. 9b). Strikingly, we observed that most of the immune-signaling-related downregulated genes are directly or indirectly connected to the viral proteins (Fig. 9b); this suggests the probable roles of the virus in the differential host responses in the COVID-19 affected patients.

Furthermore, to understand if there are any viral factor dependent enhancement of integrin expression, we sought to establish the links between the viral proteins with integrin signaling associated genes by constructing a functional network with the viral-host protein–protein interaction data with the highly upregulated genes observed in lungs (from the comparison analysis between the lung and nasopharyngeal samples) (Fig. 9c). From this constructed network, we observed that viral proteins such as ORF10, N, ORF9b, NSP7, NSP15, NSP5, M, NSP13, NSP2, NSP9, ORF8, ORF9c, NSP12, and NSP1 can directly or indirectly interact with the differentially expressed genes in lungs (Fig. 9c), suggesting the putative mechanism behind the deregulated integrin signaling to promote the viral invasion in lungs (Fig. 10).

Discussion

For a better understanding of the host-virus interaction in the SARS-CoV-2 pathogenesis, transcriptional responses of hosts play an enormous role. In this context, we aimed to discover the host transcriptome response upon SARS-CoV-2 infection by performing and analyzing total RNA-seq from the nasopharyngeal samples of four COVID-19 positive individuals. Moreover, we compared the transcriptome from different SARS-CoV-2 infection models, particularly, we compared the differential gene expression of the lung biopsy samples with the nasopharyngeal samples of ours to illustrate the possible molecular mechanisms behind the lung damages in severe COVID-19 patients.

Previously, host transcriptional responses reported by Blanco-melo et al. [24] and Butler et al. [25] suggested a potential increase in the host antiviral immune responses such as interferon signaling, interferon stimulated gene signaling, chemokine signaling, and cytokine signaling; however, Blanco-melo et al. [61] also reported the presence of low IFN-I and IFN-III in COVID-19 patient's lung cells. We observed similar host immune responses,


a
sd
rank(mean)
0
20000
40000
60000
count
403
55
7
1
b
Color Key and Histogram
Count
Value
lung : wt
lung : treat
nasal : treat
nasal : wt
c
PC1: 65% variance
PC2: 25% variance
condition
lung
nasal
type
treat
wt
d
type
condition
RN7SL2
RN7SL2
RN7SL1
RN7SL1
RNA5SP202
FP236383.9
FP671120.6
RNA5SP389
RNA5SP370
RNU1-2
RNVU1-18
RNU1-1
RNU1-3
RNA5S12
RNA5S17
RNA5S11
RNA5S14
RNA5S16
RNA5S13
RNA5S10
RNA5S15
RNA5S8
RNA5S1
RNA5S7
RNA5S3
RNA5S2
RNA5S6
RNA5S5
RNA5S4
MT-ND5
MT-CO1
MT-ND4
MALAT1
NEAT1
AL162581.1
RNA5-8SN1
FP671120.4
RNA5-8SN2
RNA5-8SN3
RN7SL5P
RN7SL4P
MT-RNR2
MT-RNR1
FP236383.10
FP671120.4
FP236383.3
RNA28S5
FP236383.12
FP671120.7
FP236383.4
nasal_wt1
nasal_wt2
nasal_wt3
nasal_wt4
nasal_S2
nasal_S3
nasal_S4
nasal_S9
lung_wt1
lung_wt2
lung_wt3
lung_wt4
lung_wt5
Lung_GSE150316_GSM4546576
Lung_GSE150316_GSM4546581
Lung_GSE150316_GSM4546586
Lung_GSE150316_GSM4546589
Lung_GSE150316_GSM4546596
0 5 10 15 20
type
treat
wt
condition
lung
nasal
e
Tagwise
Common
Trend
Biological coefficient of variation
Average log CPM
f
1*A −1*B −1*C 1*D
NotSig
Up
Down
log−fold−change
Average log CPM
g
ADAM15
CD14
CD47
CD81
CD9
CDH1
COL6A3
CTNNB1
ERBB3
EXOC1
EXOC3
F11R
GSN
HMGB1
ITGA6
ITGAV
LAMB3
LAMC1
MST1R
SDC1
YWHAB
COL5A1
Log2 (Fold change) scale
-2
0
2

(See figure on next page.)

Fig. 8 Enrichment analysis and comparison between deregulated genes and the genes of some selected processes in SARS-CoV-2 infected nasopharyngeal samples versus SARS-CoV-2 infected lung biopsy samples, using **a** GOBP module, **b** KEGG pathway, **c** Bioplanet pathway module, **d** Reactome pathway module. Selected significant terms are represented in heatmaps. Color schemes are similar to Fig. 3. For individual processes, blue means presence (significantly differentially expressed gene) while grey means absence (not significantly differentially expressed genes for this module for this experimental condition). Here, Up and down denote Upregulated and Downregulated, respectively

interferon, and cytokine signaling in our reported COVID-19 patients too. Moreover, we also observed a stimulated innate immune response in our patients which was also reported for other COVID-19 patients [62].

Astoundingly, important signaling pathways those elicit antiviral immune responses such as apoptosis [20], phagosome formation [63], antigen processing and presentation [64], Natural killer cell mediated cytotoxicity [65], and Toll-like receptor signaling [66] were found downregulated in these COVID-19 patients. Also, pathways such as HIF-1 response [67], PI3K-Akt signaling [68], and IL-17 signaling [69] were also found deregulated, which could assist the COVID-19 patients suffering from hypoxia, lung injury, and inflammation of the respiratory tract.

All of our patients showed dysosmia which is also a commonly observed features in most other COVID-19 patients around the world. This might have occurred due to the hypothesized reasons reported by Breguglio et al. [70]. Interestingly, our patients' nasopharyngeal data also provides supportive clues such as overexpressing local cytokine signaling, inflammatory responses and accumulation of innate immune cells in the nasopharyngeal regions; all of which might contribute towards the destabilization of olfaction within these patients.

While we were comparing the nasopharyngeal cell's transcriptional responses with other SARS-CoV-2 infection models, we observed that lung cells elicited the immense cytokine and inflammatory responses against the invading viral pathogen. These overstimulated responses sometimes can do irreversible damages to the lungs [71]. This might shed insights into the COVID-19 disease severity when the viral infection progresses into the lungs.

Though an increased amount of ACE2 will facilitate the invasion of SARS-CoV-2, nonetheless, we observed a significant downregulation of ACE2 in lung cells; Hou et al. reported similar phenomenon in an earlier study [72]. This phenomenon could backup the concept of ACE2 downregulation by SARS-CoV-2 itself after using it [73], thus reducing the organ protective roles of ACE2 [74] and resulting in progressive lung damages.

Integrins were reported important for the entry of SARS-CoV into the host cells [39], so it was speculated similar phenomenon might also be present in SARS-CoV-2. This idea is further intensified after the study by Sigrist et al. [37], who suggested the presence of an integrin-binding RGD motif in the spike of SARS-CoV-2. Surprisingly, upon the gene expression comparison between the different SARS-CoV-2 infected cells, we observed several folds upregulated expressions of genes encoding integrins in lung cells. This observation could support the idea of increased viral infections in lungs might be happening due to the overexpression of these probable attachment proteins. Also, the network analysis suggests a probable mechanism of upregulation of these proteins by the virus itself by the putative interactions through its proteins. As our study is based on the data acquired from a limited number of samples, therefore, more targeted studies with a larger sample size should be undertaken for conclusive evidence supporting this phenomenon.

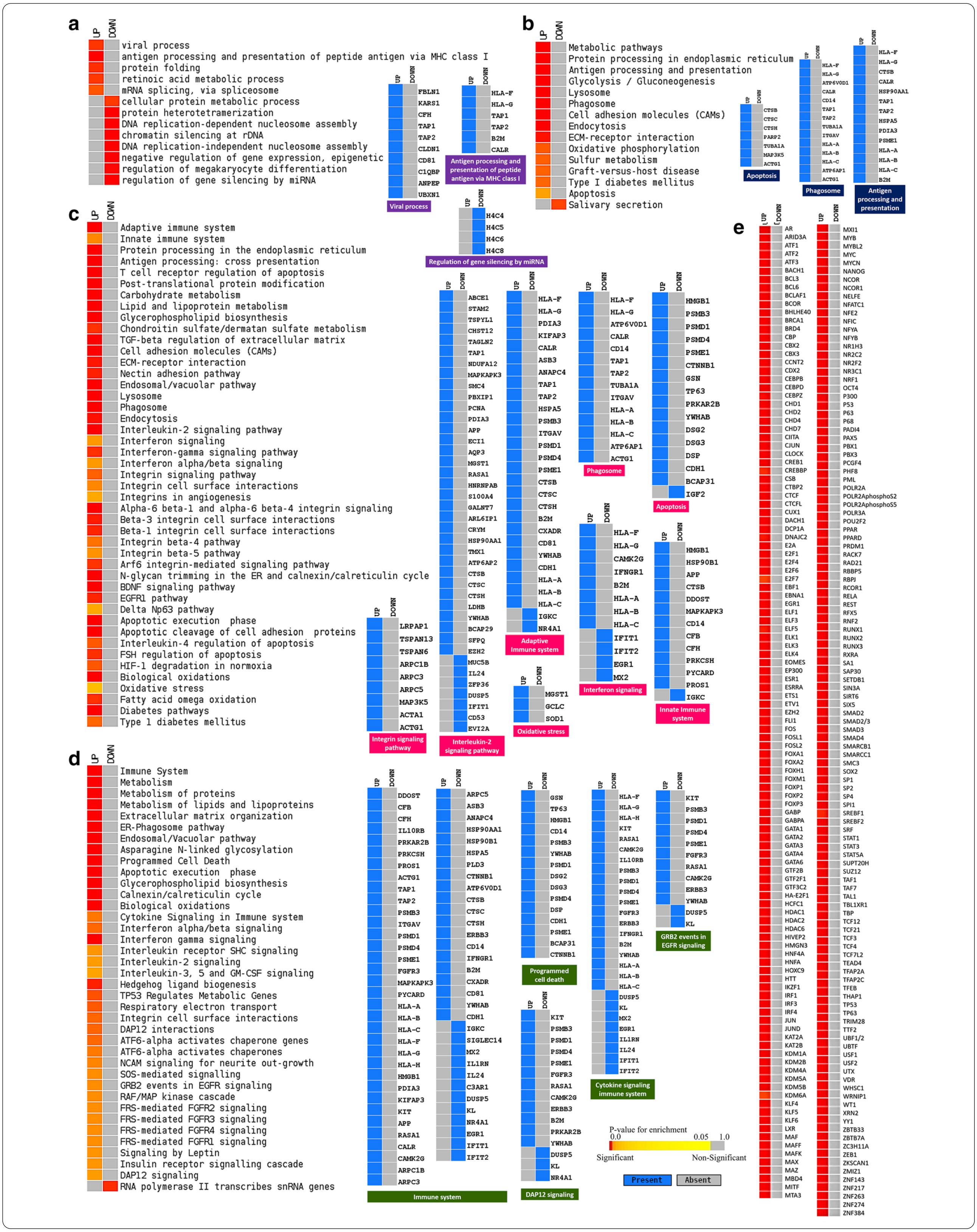
Conclusion

In this study, we present the very first report of the host transcriptional response data from COVID-19 patients of the South-Asian region along with the SARS-CoV-2 isolates obtained from these patients. This data might provide newer insights into the host responses against the virus in the different parts of the respiratory tract. However, a limited number of patient data is used here, but subsequent incorporation of more patient data from other parts of the world will significantly increase the understanding of this complex host-virus response in COVID-19, which will help in designing therapeutic interventions as well as in current clinical management of the patients.

Methods

Sample collection and virus detection by Real-time reverse transcription-quantitative PCR (RT-qPCR)

The nasopharyngeal swab samples were collected from patients suspicious of COVID-19 and placed in sample collection vial containing normal saline. Collected samples were preserved at − 20 °C until further use for RNA extraction and RT-qPCR assay. The RT-qPCR was performed for ORF1ab and N genes of SARS-CoV-2 using Novel Coronavirus (2019-nCoV) Nucleic Acid Diagnostic Kit (PCR-Fluorescence Probing) of Sansure Biotech Inc. according to the manufacturer's instructions. RNA was extracted from a 20 μL swab sample through lysis with sample release reagent provided by the kit and then directly used for RT-qPCR. Thermal cycling was


a
UP
DOWN
viral process
antigen processing and presentation of peptide antigen via MHC class I
protein folding
retinoic acid metabolic process
mRNA splicing, via spliceosome
cellular protein metabolic process
protein heterotetramerization
DNA replication-dependent nucleosome assembly
chromatin silencing at rDNA
DNA replication-independent nucleosome assembly
negative regulation of gene expression, epigenetic
regulation of megakaryocyte differentiation
regulation of gene silencing by miRNA
FBLN1
KARS1
CFH
TAP1
TAP2
CLDN1
CD81
C1QBP
ANPEP
UBXN1
Viral process
HLA-F
HLA-G
TAP1
TAP2
B2M
CALR
Antigen processing and presentation of peptide antigen via MHC class I
H4C4
H4C5
H4C6
H4C8
Regulation of gene silencing by miRNA
b
Metabolic pathways
Protein processing in endoplasmic reticulum
Antigen processing and presentation
Glycolysis / Gluconeogenesis
Lysosome
Phagosome
Cell adhesion molecules (CAMs)
Endocytosis
ECM-receptor interaction
Oxidative phosphorylation
Sulfur metabolism
Graft-versus-host disease
Type I diabetes mellitus
Apoptosis
Salivary secretion
CTSB
CTSC
CTSH
PARP2
TUBA1A
MAP3K5
ACTG1
Apoptosis
HLA-F
HLA-G
ATP6V0D1
CALR
CD14
TAP1
TAP2
TUBA1A
ITGAV
HLA-A
HLA-B
HLA-C
ATP6AP1
ACTG1
Phagosome
HLA-F
HLA-G
CTSB
CALR
HSP90AA1
TAP1
TAP2
HSPA5
PDIA3
PSME1
HLA-A
HLA-B
HLA-C
B2M
Antigen processing and presentation
c
Adaptive immune system
Innate immune system
Protein processing in the endoplasmic reticulum
Antigen processing: cross presentation
T cell receptor regulation of apoptosis
Post-translational protein modification
Carbohydrate metabolism
Lipid and lipoprotein metabolism
Glycerophospholipid biosynthesis
Chondroitin sulfate/dermatan sulfate metabolism
TGF-beta regulation of extracellular matrix
Cell adhesion molecules (CAMs)
ECM-receptor interaction
Nectin adhesion pathway
Endosomal/vacuolar pathway
Lysosome
Phagosome
Endocytosis
Interleukin-2 signaling pathway
Interferon signaling
Interferon-gamma signaling pathway
Interferon alpha/beta signaling
Integrin signaling pathway
Integrin cell surface interactions
Integrins in angiogenesis
Alpha-6 beta-1 and alpha-6 beta-4 integrin signaling
Beta-3 integrin cell surface interactions
Beta-1 integrin cell surface interactions
Integrin beta-4 pathway
Integrin beta-5 pathway
Arf6 integrin-mediated signaling pathway
N-glycan trimming in the ER and calnexin/calreticulin cycle
BDNF signaling pathway
EGFR1 pathway
Delta Np63 pathway
Apoptotic execution phase
Apoptotic cleavage of cell adhesion proteins
Interleukin-4 regulation of apoptosis
FSH regulation of apoptosis
HIF-1 degradation in normoxia
Biological oxidations
Oxidative stress
Fatty acid omega oxidation
Diabetes pathways
Type 1 diabetes mellitus
LRPAP1
TSPAN13
TSPAN6
ARPC1B
ARPC3
ARPC5
MAP3K5
ACTA1
ACTG1
Integrin signaling pathway
ABCE1
STAM2
TSPYL1
CHST12
TAGLN2
TAP1
NDUFA12
MAPKAPK3
SMC4
PBXIP1
PCNA
PDIA3
APP
ECI1
AQP3
MGST1
RASA1
HNRNPAB
S100A4
GALNT7
ARL6IP1
CRYM
HSP90AA1
TMX1
ATP6AP2
CTSB
CTSC
CTSH
LDHB
YWHAB
BCAP29
SFPQ
EZH2
MUC5B
IL24
ZFP36
DUSP5
IFIT1
CD53
EVI2A
Interleukin-2 signaling pathway
HLA-F
HLA-G
PDIA3
KIFAP3
CALR
ASB3
ANAPC4
TAP1
TAP2
HSPA5
PSMB3
ITGAV
PSMD1
PSMD4
PSME1
CTSB
CTSC
CTSH
B2M
CXADR
CD81
YWHAB
CDH1
HLA-A
HLA-B
HLA-C
IGKC
NR4A1
Adaptive Immune system
MGST1
GCLC
SOD1
Oxidative stress
HLA-F
HLA-G
ATP6V0D1
CALR
CD14
TAP1
TAP2
TUBA1A
ITGAV
HLA-A
HLA-B
HLA-C
ATP6AP1
ACTG1
Phagosome
HLA-F
HLA-G
CAMK2G
IFNGR1
B2M
HLA-A
HLA-B
HLA-C
IFIT1
IFIT2
EGR1
MX2
Interferon signaling
HMGB1
PSMB3
PSMD1
PSMD4
PSME1
CTNNB1
GSN
TP63
PRKAR2B
YWHAB
DSG2
DSG3
DSP
CDH1
BCAP31
IGF2
Apoptosis
HMGB1
HSP90B1
APP
CTSB
DDOST
MAPKAPK3
CD14
CFB
CFH
PRKCSH
PYCARD
PROS1
IGKC
Innate Immune system
d
Immune System
Metabolism
Metabolism of proteins
Metabolism of lipids and lipoproteins
Extracellular matrix organization
ER-Phagosome pathway
Endosomal/Vacuolar pathway
Asparagine N-linked glycosylation
Programmed Cell Death
Apoptotic execution phase
Glycerophospholipid biosynthesis
Calnexin/calreticulin cycle
Biological oxidations
Cytokine Signaling in Immune system
Interferon alpha/beta signaling
Interferon gamma signaling
Interleukin receptor SHC signaling
Interleukin-2 signaling
Interleukin-3, 5 and GM-CSF signaling
Hedgehog ligand biogenesis
TP53 Regulates Metabolic Genes
Respiratory electron transport
Integrin cell surface interactions
DAP12 interactions
ATF6-alpha activates chaperone genes
ATF6-alpha activates chaperones
NCAM signaling for neurite out-growth
SOS-mediated signalling
GRB2 events in EGFR signaling
RAF/MAP kinase cascade
FRS-mediated FGFR2 signaling
FRS-mediated FGFR3 signaling
FRS-mediated FGFR4 signaling
FRS-mediated FGFR1 signaling
Signaling by Leptin
Insulin receptor signalling cascade
DAP12 signaling
RNA polymerase II transcribes snRNA genes
DDOST
CFB
CFH
IL10RB
PRKAR2B
PRKCSH
PROS1
ACTG1
TAP1
TAP2
PSMB3
ITGAV
PSMD1
PSMD4
PSME1
FGFR3
MAPKAPK3
PYCARD
HLA-A
HLA-B
HLA-C
HLA-F
HLA-G
HLA-H
HMGB1
PDIA3
KIFAP3
KIT
APP
RASA1
CALR
CAMK2G
ARPC1B
ARPC3
ARPC5
ASB3
ANAPC4
HSP90AA1
HSP90B1
HSPA5
PLD3
CTNNB1
ATP6V0D1
CTSB
CTSC
CTSH
ERBB3
CD14
IFNGR1
B2M
CXADR
CD81
YWHAB
CDH1
IGKC
SIGLEC14
MX2
IL1RN
IL24
C3AR1
DUSP5
KL
NR4A1
EGR1
IFIT1
IFIT2
Immune system
GSN
TP63
HMGB1
CD14
PSMB3
YWHAB
PSMD1
DSG2
DSG3
PSMD4
DSP
CDH1
PSME1
BCAP31
CTNNB1
Programmed cell death
KIT
PSMB3
PSMD1
PSMD4
PSME1
FGFR3
RASA1
CAMK2G
ERBB3
B2M
PRKAR2B
YWHAB
DUSP5
KL
NR4A1
DAP12 signaling
HLA-F
HLA-G
HLA-H
KIT
RASA1
CAMK2G
IL10RB
PSMB3
PSMD1
PSMD4
PSME1
FGFR3
ERBB3
IFNGR1
B2M
YWHAB
HLA-A
HLA-B
HLA-C
DUSP5
KL
MX2
EGR1
IL1RN
IL24
IFIT1
IFIT2
Cytokine signaling immune system
KIT
PSMB3
PSMD1
PSMD4
PSME1
FGFR3
RASA1
CAMK2G
ERBB3
YWHAB
DUSP5
KL
GRB2 events in EGFR signaling
P-value for enrichment
0.0
0.05
1.0
Significant
Non-Significant
Present
Absent
e
AR
ARID3A
ATF1
ATF2
ATF3
BACH1
BCL3
BCL6
BCLAF1
BCOR
BHLHE40
BRCA1
BRD4
CBP
CBX2
CBX3
CCNT2
CDX2
CEBPB
CEBPD
CEBPZ
CHD1
CHD2
CHD4
CHD7
CIITA
CJUN
CLOCK
CREB1
CREBBP
CSB
CTBP2
CTCF
CTCFL
CUX1
DACH1
DCP1A
DNAJC2
E2A
E2F1
E2F4
E2F6
E2F7
EBF1
EBNA1
EGR1
ELF1
ELF3
ELF5
ELK1
ELK3
ELK4
EOMES
EP300
ESR1
ESRRA
ETS1
ETV1
EZH2
FLI1
FOS
FOSL1
FOSL2
FOXA1
FOXA2
FOXH1
FOXM1
FOXP1
FOXP2
FOXP3
GABP
GABPA
GATA1
GATA2
GATA3
GATA4
GATA6
GTF2B
GTF2F1
GTF3C2
HA-E2F1
HCFC1
HDAC1
HDAC2
HDAC6
HIVEP2
HMGN3
HNF4A
HNFA
HOXC9
HTT
IKZF1
IRF1
IRF3
IRF4
JUN
JUND
KAT2A
KAT2B
KDM1A
KDM2B
KDM4A
KDM5A
KDM5B
KDM6A
KLF4
KLF5
KLF6
LXR
MAF
MAFF
MAFK
MAX
MAZ
MBD4
MITF
MTA3
MXI1
MYB
MYBL2
MYC
MYCN
NANOG
NCOR
NCOR1
NELFE
NFATC1
NFE2
NFIC
NFYA
NFYB
NR1H3
NR2C2
NR2F2
NR3C1
NRF1
OCT4
P300
P53
P63
P68
PADI4
PAX5
PBX1
PBX3
PCGF4
PHF8
PML
POLR2A
POLR2AphosphoS2
POLR2AphosphoS5
POLR3A
POU2F2
PPAR
PPARD
PRDM1
RACK7
RAD21
RBBP5
RBPJ
RCOR1
RELA
REST
RFX5
RNF2
RUNX1
RUNX2
RUNX3
RXRA
SA1
SAP30
SETDB1
SIN3A
SIRT6
SIX5
SMAD2
SMAD2/3
SMAD3
SMAD4
SMARCB1
SMARCC1
SMC3
SOX2
SP1
SP2
SP4
SPI1
SREBF1
SREBF2
SRF
STAT1
STAT3
STAT5A
SUPT20H
SUZ12
TAF1
TAF7
TAL1
TBL1XR1
TBP
TCF12
TCF21
TCF3
TCF4
TCF7L2
TEAD4
TFAP2A
TFAP2C
TFEB
THAP1
TP53
TP63
TRIM28
TTF2
UBF1/2
UBTF
USF1
USF2
UTX
VDR
WHSC1
WRNIP1
WT1
XRN2
YY1
ZBTB33
ZBTB7A
ZC3H11A
ZEB1
ZKSCAN1
ZMIZ1
ZNF143
ZNF217
ZNF263
ZNF274
ZNF384

(See figure on next page.)

Fig. 9 Interactions between SARS-CoV-2 proteins and differentially expressed genes of host. **a** Venn diagram showing the commonly deregulated genes between deregulated genes in our nasopharyngeal samples and Gordon et al. reported viral protein-interacting high confidence host proteins. Network representing the interactions between genes in **b**. Deregulated genes in nasopharyngeal samples along with SARS-CoV-2 proteins and Gordon et al. described viral interacting host proteins, and **c** Differentially expressed Integrin related genes in lungs compared to the nasal samples along with SARS-CoV-2 proteins and Gordon et al. described viral interacting host proteins. Hexagon, ellipse, rounded rectangle represents viral proteins, process-related genes, and proteins that interact with viral proteins, respectively. Expression values of the genes and represented in a color-coded scale. Color towards red indicating higher expression, while color towards green indicating little to no expression

performed at 50 °C for 30 min for reverse transcription, followed by 95 °C for 1 min and then 45 cycles of 95 °C for 15 s, 60 °C for 30 s on an Analytik-Jena qTOWER instrument (Analytik Jena, Germany).

RNA sequencing

Total RNA was extracted from nasopharyngeal swab samples (labeled as S2, S3, S4, S9) collected from SARS-COV-2 infected COVID-19 patients using TRIzol (Invitrogen) reagent following the manufacturer's protocol. RNA-seq libraries were prepared from total RNA using TruSeq Stranded Total RNA Library Prep kit (Illumina) according to the manufacturer's instructions where the first-strand cDNA was synthesized using SuperScript II Reverse Transcriptase (Thermo Fisher) and random primers. Paired-end (150 bpreads) sequencing of the RNA library was performed on the Illumina NextSeq 500 platform.

Data processing and identification of the viral agent

Firstly, the sequencing reads were adapter and quality trimmed using the Trimmomatic program [75]. The remaining reads were mapped against the SARS-CoV-2 reference sequence (NC_045512.2) using Bowtie 2 [76]. Then the mapped reads were assembled de novo using Megahit (v.1.1.3) [77].

Mapping of the RNA-seq reads onto SARS-CoV-2 reference genome

We mapped the normalized (by count per million mapped reads-CPM) RNA-seq reads onto the SARS-CoV-2 genome track of the UCSC genome browser [78] using the "bamCoverage" feature of deepTools2 suite [79].

Identification of SARS-CoV-2 genome variations and variation annotation

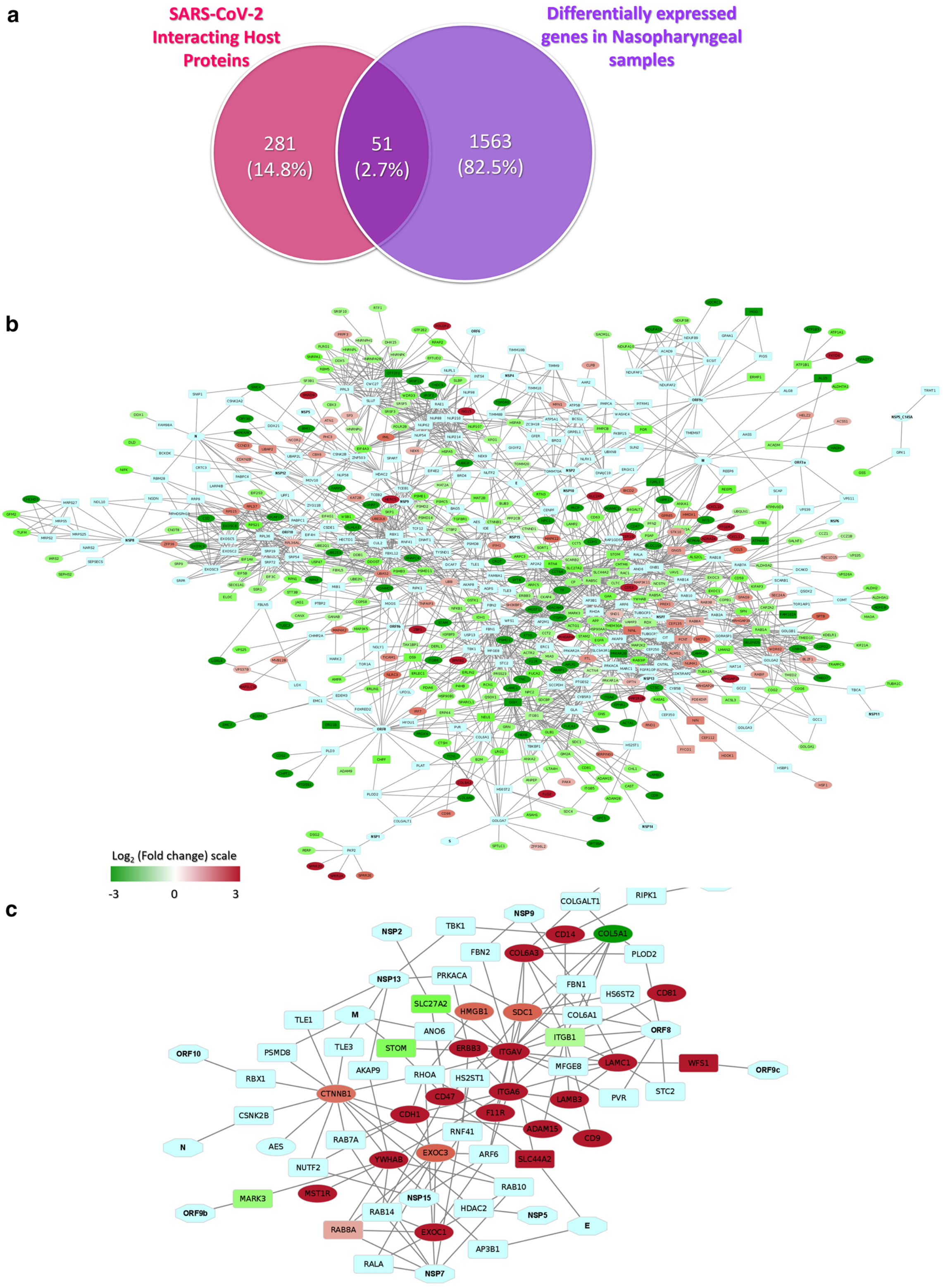
We identified the variations within our sequenced SARS-CoV-2 genome using the "Variation Identification" (https://bigd.big.ac.cn/ncov/online/tool/variation) tool of "2019 Novel Coronavirus Resource (2019nCoVR)" portal of China National Center for Bioinformation [80]. We then annotated the variations of the isolated SARS-CoV-2 isolates using the "Variation Annotation" (https://bigd.big.ac.cn/ncov/online/tool/annotation) tool from the same portal [80]. We also gathered the global frequency of every identified variation using this same information portal [80]. Different representations showing the information regarding the variations were produced using the Microsoft Excel program [81]. The impacts of the variations were further characterized utilizing the Ensembl Variant Effect Predictor (VEP) tool [82].

Analysis of RNA-seq expression data

We analyzed both our RNA-seq and some publicly available RNA-seq data for COVID-19 host transcriptional profile analysis. Publicly available Illumina sequenced RNA-seq raw FastQ reads were extracted from the GEO database (accessions of the data used can be found in Additional file 11) [83]. We have checked the raw sequence quality using FastQC program (v0.11.9) [84] and found that the "Per base sequence quality", and "Per sequence quality scores" were high over the threshold for all sequences (Additional file 12). The mapping of reads was done with TopHat (tophat v2.1.1 with Bowtie v2.4.1) [85]. Short reads were uniquely aligned allowing at best two mismatches to the human reference genome from (GRCh38) as downloaded from the UCSC database [86]. Sequence matched exactly more than one place with equally quality were discarded to avoid bias [87]. The reads that were not mapped to the genome were utilized to map against the transcriptome (junctions mapping). Ensembl gene model [88] (version 99, as extracted from UCSC) was used for this process. After mapping, we used the SubRead package featureCount (v2.21) [89] to calculate absolute read abundance (read count, rc) for each transcript/gene associated to the Ensembl genes.

Differential gene expression analysis

To obtain the differential gene expression profile of our studied nasal samples, we utilized the the RNA-seq data recorded from nasal epithelial cells of 4 different non-asthmatic adult individuals as normal controls (Additional file 11); these cells were taken 7 days before the original infection analysis (GEO accession: GSE97668). For the differential gene expression analysis of COVID-19 affected lungs, we've taken the RNA-seq data from the lung biopsy of a deceased COVID-19 patient and


a
SARS-CoV-2 Interacting Host Proteins
Differentially expressed genes in Nasopharyngeal samples
281 (14.8%)
51 (2.7%)
1563 (82.5%)
b
Log2 (Fold change) scale
-3
0
3
c
NSP2
TBK1
NSP9
COLGALT1
RIPK1
CD14
COL5A1
FBN2
COL6A3
PLOD2
NSP13
PRKACA
FBN1
HS6ST2
CD81
SLC27A2
HMGB1
SDC1
COL6A1
TLE1
M
ANO6
ORF8
ITGB1
ORF10
PSMD8
TLE3
STOM
ERBB3
ITGAV
MFGE8
LAMC1
WFS1
ORF9c
AKAP9
RHOA
HS2ST1
RBX1
CTNNB1
CD47
ITGA6
LAMB3
PVR
STC2
CDH1
F11R
CSNK2B
RNF41
ADAM15
CD9
N
AES
RAB7A
EXOC3
ARF6
SLC44A2
NUTF2
YWHAB
RAB10
MARK3
MST1R
NSP15
HDAC2
ORF9b
RAB14
NSP5
E
RAB8A
EXOC1
AP3B1
RALA
NSP7

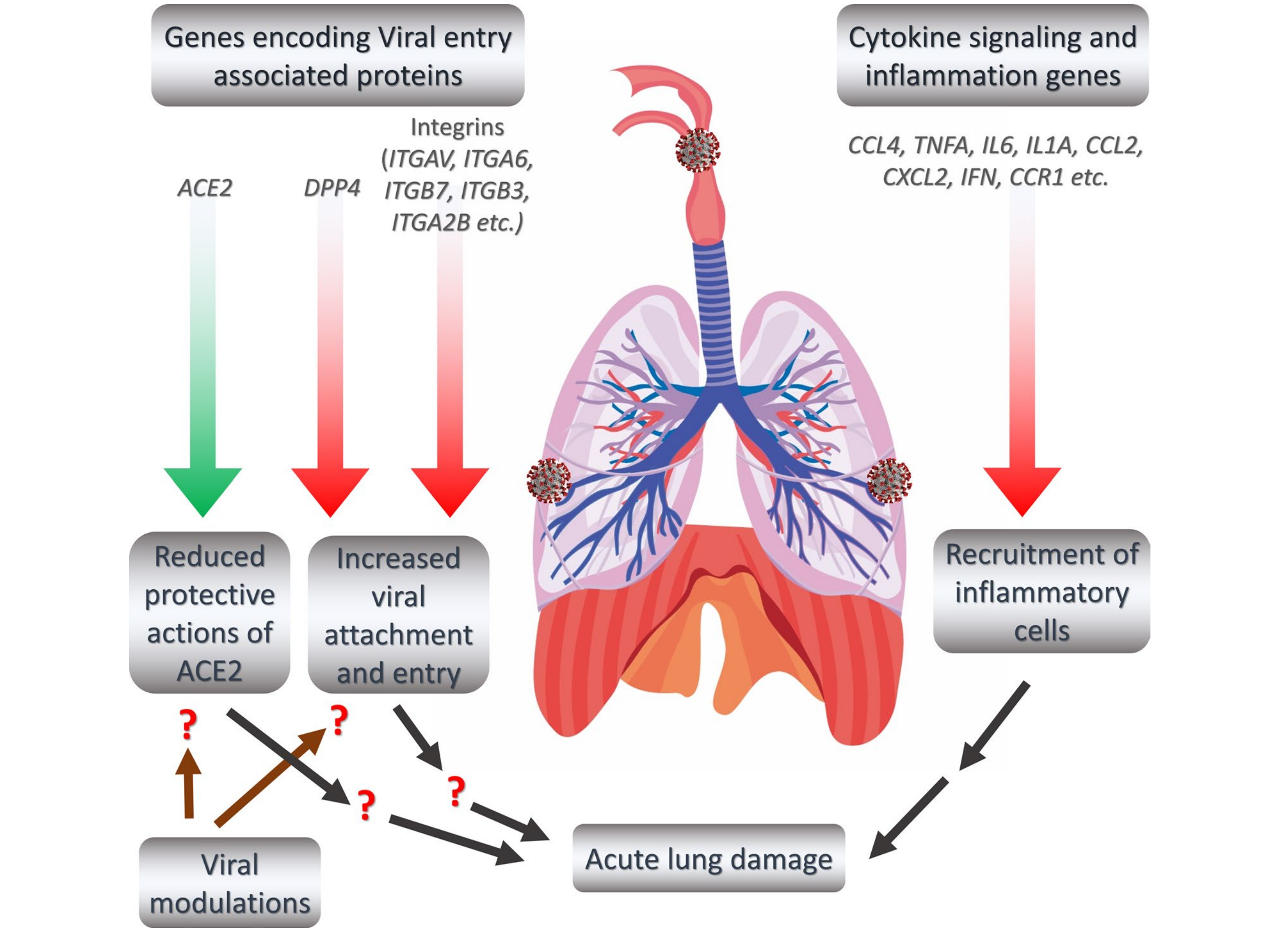

Genes encoding Viral entry associated proteins
Cytokine signaling and inflammation genes
ACE2
DPP4
Integrins (ITGAV, ITGA6, ITGB7, ITGB3, ITGA2B etc.)
CCL4, TNFA, IL6, IL1A, CCL2, CXCL2, IFN, CCR1 etc.
Reduced protective actions of ACE2
Increased viral attachment and entry
Recruitment of inflammatory cells
?
?
?
?
Viral modulations
Acute lung damage

Fig. 10 Schematic representation of putative mechanisms of acute lung damages in COVID-19. Red arrow suggesting the increasing expression values, while the green arrow indicating the decreasing expression from the nasopharyngeal region to the lung

the associated controls from the original study (GEO accession: GSE147507); and another set of data from five deceased COVID-19 patient's (initially all of them were hospitalized) lung autopsy and associated controls from the original study (GEO accession: GSE150316) (Additional file 11). Moreover, for the differential transcriptome documented for the cell lines, we used the RNA-seq data from infected cell lines and associated controls from GEO datasets GSE148729 and GSE147507 (Additional file 11).

For differential expression (DE) analysis, we used DESeq2 (v1.26.0) [90] with R (v3.6.2; 2019–07-05) that uses a model based on the negative binomial distribution. To avoid false positive, we considered only those transcripts where at least 10 reads are annotated in at least one of the samples used in this study and also applied a minimum Log2 fold change of 0.5 for to be differentially apart from adjusted p-value cut-off of ≤ 0.05 by FDR. To assess the fidelity of the RNA-seq data used in this study and normalization method applied here, we checked the normalized Log2 expression data quality using R/Bioconductor package "arrayQualityMetrics (v3.44.0)" [91]. From these analyses, no outlier was detected in our data by "Distance between arrays", "Boxplots", and "MA plots" methods and replicate samples are clustered together (data not shown). We considered the genes upregulated which have a positive Log2 fold change value higher than 0.5, and those with a Log2 fold change value lower than − 0.5 were considered downregulated.

We also performed a multifactorial differential gene expression analysis using the edgeR tool [44] following the generalized linear model (GLM) experimental design- log2 (lung samples/normal lung control samples)/ log2 (our studied Nasal samples/normal nasal control samples); we used the autopsy samples of COVID-19 patients and associated controls from (GEO accession: GSE150316) as lung sample & controls, and we used our analyzed nasal COVID-19 transcriptomes as nasal samples alongwith the RNA-seq data from (GEO accession: GSE97668) as normal controls.

Construction of phylogenetic tree

We constructed a Neighbour-Joining phylogenetic tree with all available 145 SARS-CoV-2 genomes of Bangladeshi isolates (retrieved on 6th May from GISAID [92]). Firstly, the genome sequences were aligned using MAFFT [93] tool using the auto-configuration. Then we used MEGA X [94] for constructing the phylogenetic tree utilizing 500 bootstrapping with substitution model/method: maximum composite likelihood, uniform rates of variation among sites, the partial deletion of gaps/missing data and site coverage cutoff 95%.

Functional enrichment analysis

We utilized Gitools (v1.8.4) for enrichment analysis and heatmap generation [95]. We have utilized the Gene Ontology Biological Processes (GOBP) [96], Bioplanet pathways [97], KEGG pathway [98], and Reactome pathway [99] modules for the overrepresentation analysis. Resulting p-values were adjusted for multiple testing using the Benjamin and Hochberg's method of False Discovery Rate (FDR) [100].

Retrieval of the host proteins that interact with SARS-CoV-2

We have obtained the list of human proteins that form high confidence interactions with SARS-CoV-2 proteins from conducted previously study [21] and processed their provided protein names into the associated HGNC official gene symbols.

Construction of biological networks

Construction, visualization, and analysis of biological networks with differentially expressed genes, their associated transcription factors, and interacting viral proteins were executed in the Cytoscape software (v3.8.0) [101]. We used the STRING [102] database to extract the highest confidences (0.9) edges only for the protein–protein interactions to reduce any false positive connection.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-020-02695-0.

Additional file 1. Isolate-wise variation information of the four SARS-CoV-2 isolates used in this study.

Additional file 2: Figure S1. A. Snapshot of Nextstrain data portal showing the phylogenetic relationship of two SARS-CoV-2 isolates used in this study. Isolates of this study are indicated using a red arrow. **B.** Phylogenetic tree of Bangladeshi SARS-CoV-2 isolates. Neighbor-joining tree using MEGA tools. Isolates reported in this study are indicated with a red arrow. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.01403419 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. This analysis involved 145 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 29827 positions in the final dataset. Values represent bootstrap numbers (%).

Additional file 3. Patient specific clinical features observed in the analyzed four COVID-19 patients.

Additional file 4. Differentially expressed genes found in the four nasal samples of COVID-19 patients.

Additional file 5: Figure S2. A. Hierarchically clustered heatmap representing the patient-wise complete expression profiles. Normalized Log_2 fold changes compared to average normal expression values across the samples are represented in a color coded heatmap, and for one of the four samples only protein coding genes (with Log_2 fold change > 0.5) are represented. Pearson correlation distance was utilized for this hierarchical clustering of the genes. **B.** Deregulated genes of selected terms from Fig. 3 in different SARS-CoV-2 infection systems. Genes of selected significant terms are represented here. For individual processes, blue means presence (differentially expressed gene of the module term) while grey means absence (not differentially expressed in the experimental condition in that module term). Processes in the green, blue, red color background represent KEGG, Bioplanet, GOBP enriched terms, respectively.

Additional file 6. Differentially expressed genes in different SARS-CoV-2 infected cell types.

Additional file 7: Figure S3. Deregulated genes of selected terms from Fig. 3 in different SARS-CoV-2 infection systems. For individual processes, blue means presence (differentially expressed gene of the module term) while grey means absence (not differentially expressed in the experimental condition in that module term). Processes in the green, blue, red color background represent KEGG, Bioplanet, GOBP enriched terms, respectively.

Additional file 8. Genes and associated terms used for filtering the expression values used in Fig. 6.

Additional file 9: Figure S4. Expanded view of the heatmaps A, B, C of Fig. 6.

Additional file 10. Differentially expressed genes in SARS-CoV-2 infected lungs compared to the our nasal samples used in this study.

Additional file 11. Sources of the data used in this study.

Additional file 12. Per base sequence quality reports of the generated RNA-seq reads of the four COVID-19 infected nasal samples used in this study.

Acknowledgements

We would like to thank all the authors who have kindly deposited and shared genome data on GISAID (https://www.gisaid.org/). We would also want to convey our gratitude to the developers maintaining the NextStrain novel coronavirus data portal (https://nextstrain.org/).

Authors' contributions

ABMMKI designed the workflow and conceived the project. AMAMZS performed sample collection and detection. RA, MSH, SMTK, and MSI performed the RNA sequencing and viral genome assembly. ABMMKI and MAAKK performed the RNA-seq data analysis, comparative genomics, and other bioinformatic analyses. MAAKK and ABMMKI wrote the manuscript. ABMMKI and MAAKK contributed equally to this work. All authors read and approved the final manuscript.

Funding

This project was not associated with any internal or external source of funding.

Availability of data and materials

SARS-CoV-2 genomes are deposited at GISAID initiatives portal (https://www.gisaid.org/) with accessions- EPI_ISL_450340, EPI_ISL_450341, EPI_ISL_450342, EPI_ISL_450345. Raw RNA-seq data are deposited at NCBI-GEO (https://www.ncbi.nlm.nih.gov/geo/) under the accession- GSM4667504, GSM4667505, GSM4667506, GSM4667507. Additionally, publicly available data were utilized (Additional file 11). Analyses generated data are deposited as Additional files.

Ethics approval and consent to participate

Patients' written-consents were taken before the collection of the samples. An institutional ethical clearance (Ref. No: 100/Biol. Scs., dated: 23/08/2020) was obtained from the Ethical Review Committee of the Faculty of Biological Sciences, University of Dhaka.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1] Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka 1000, Bangladesh. [2] Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh. [3] Basic and Applied Research On Jute Project, Bangladesh Jute Research Institute, Dhaka, Bangladesh. [4] Department of Pathology and Parasitology, Chittagong Veterinary and Animal Sciences University (CVASU), Khulshi, Chittagong, Bangladesh.

Received: 11 November 2020 Accepted: 29 December 2020
Published online: 07 January 2021

References

1. Worldometer. Coronavirus Cases. New York: Worldometer ; 2020. p. 1–22.
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020;395(10224):565–74.
3. NCBI-Gene. Gene Links for Nucleotide (Select 1798174254) - Gene - NCBI. 2020.
4. Jiang S, Du L, Shi Z. An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. Emerg Microb Infect. 2020;9(1):275–7.
5. Liao J, Fan S, Chen J, Wu J, Xu S, Guo Y, et al. Epidemiological and clinical characteristics of COVID-19 in adolescents and young adults. Innovation. 2020;1(1):100001.
6. Koh J, Shah SU, Chua PEY, Gui H, Pang J. Epidemiological and clinical characteristics of cases during the early phase of COVID-19 pandemic: a systematic review and meta-analysis. Front Med (Lausanne). 2020;7:295.
7. 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 The Lancet. 2020;395(10223):497–506.
8. Galiatsatos P. What Coronavirus Does to the Lungs: Johns Hopkins Medicine; 2020. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/what-coronavirus-does-to-the-lungs. Accessed 24 Apr 2020
9. Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, et al. Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan. JAMA Neurology: China; 2020.
10. Zheng Y-Y, Ma Y-T, Zhang J-Y, Xie X. COVID-19 and the cardiovascular system. Nat Rev Cardiol. 2020;17(5):259–60.
11. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med. 2020;2:58.
12. Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR test results in patients recovered from COVID-19. JAMA. 2020;323(15):1502–3.
13. Chen D, Xu W, Lei Z, Huang Z, Liu J, Gao Z, et al. Recurrence of positive SARS-CoV-2 RNA in COVID-19: a case report. Int J Infect Dis. 2020;93:297–9.
14. Kirkcaldy RD, King BA, Brooks JT. COVID-19 and postinfection immunity: limited evidence. Many Remain Ques JAMA. 2020;323(22):2245–6.
15. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet. 2020;395(10229):1033–4.
16. Yoshikawa T, Hill TE, Yoshikawa N, Popov VL, Galindo CL, Garner HR, et al. Dynamic innate immune responses of human bronchial epithelial cells to severe acute respiratory syndrome-associated coronavirus infection. PloS ONE. 2010;5(1):e8729.
17. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. J Infect. 2020;4:12.
18. Gu J, Korteweg C. Pathology and pathogenesis of severe acute respiratory syndrome. Am J Pathol. 2007;170(4):1136–47.
19. Schäfer A, Baric RS. epigenetic landscape during coronavirus infection. Pathogens. 2017;6(1):8.
20. Fung TS, Liu DX. Human coronavirus: host-pathogen interaction. Annu Rev Microbiol. 2019;73(1):529–57.
21. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020;583(7816):459–68.
22. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neandertals. BioRxiv. 2020;2020:186296.
23. Hassan SS, Choudhury PP, Basu P, Jana SS. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. Genomics. 2020;112(5):3226–37.
24. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020;181(5):1036-45.e9.
25. Butler DJ, Mozsary C, Meydan C, Danko D, Foox J, Rosiene J, et al. Shotgun transcriptome and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. BioRxiv. 2020;2020:48066.
26. Xiong Y, Liu Y, Cao L, Wang D, Guo M, Jiang A, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. Emerg Microb Infect. 2020;9(1):761–70.
27. Biswas SK, Mudi SR. Genetic variation in SARS-CoV-2 may explain variable severity of COVID-19. Med Hypotheses. 2020;143:109877.
28. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell. 2020;182(4):812–27.
29. Hu T, Zhang Y, Li L, Wang K, Chen S, Chen J, et al. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. Virology. 2009;388(2):324–34.
30. Muramatsu T, Takemoto C, Kim Y-T, Wang H, Nishii W, Terada T, et al. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. Proc Natl Acad Sci. 2016;113(46):12997–3002.
31. Huang C, Wei P, Fan K, Liu Y, Lai L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry. 2004;43:4568–74.
32. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121–3.
33. Express TF. Manpower export from Chittagong region rises in 2017 2018. https://thefinancialexpress.com.bd/economy/bangladesh/manpower-export-from-276.
34. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. Viral and host factors related to the clinical outcome of COVID-19. Nature. 2020;583(7816):437–40.
35. Townsend MJ, Kyle TK, Stanford FC. Outcomes of COVID-19: disparities in obesity and by ethnicity/race. International Journal of Obesity. 2020;44(9):1807–9.
36. Sungnak W, Huang N, Bécavin C, Berg M, Queen R, Litvinukova M, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat Med. 2020;26(5):681–7.
37. Sigrist CJ, Bridge A, Le Mercier P. A potential role for integrins in host cell entry by SARS-CoV-2. Antiviral Res. 2020;177:104759.
38. Tresoldi I, Sangiuolo CF, Manzari V, Modesti A. SARS-COV-2 and infectivity: possible increase in infectivity associated to integrin motif expression. J Med Virol. 2020;92(10):1741–2.
39. Hänel K, Stangler T, Stoldt M, Willbold D. Solution structure of the X4 protein coded by the SARS related coronavirus reveals an

immunoglobulin like fold and suggests a binding activity to integrin I domains. J Biomed Sci. 2006;13(3):281–93.
40. Colafrancesco S, Scrivo R, Barbati C, Conti F, Priori R. Targeting the immune system for pulmonary inflammation and cardiovascular complications in COVID-19 patients. Front Immunol. 2020;11:1439.
41. Pedersen SF, Ho YC. SARS-CoV-2: a storm is raging. J Clin Investig. 2020;130(5):2202–5.
42. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int J Oral Sci. 2020;12(1):8.
43. Kuba K, Imai Y, Rao S, Gao H, Guo F, Guan B, et al. A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. Nat Med. 2005;11(8):875–9.
44. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
45. Sheppard D. Modulation of acute lung injury by integrins. Proc Am Thoracic Soc. 2012;9(3):126–9.
46. Sun C, Beard RS Jr, McLean DL, Rigor RR, Konia T, Wu MH, et al. ADAM15 deficiency attenuates pulmonary hyperpermeability and acute lung injury in lipopolysaccharide-treated mice. Am J Physiol Lung Cell Mol Physiol. 2013;304(3):L135–42.
47. Parimon T, Yao C, Habiel DM, Ge L, Bora SA, Brauer R, et al. Syndecan-1 promotes lung fibrosis by regulating epithelial reprogramming through extracellular vesicles. JCI Insight. 2019;5(17):e129359.
48. Anas A, van der Poll T, de Vos AF. Role of CD14 in lung inflammation and infection. Crit Care. 2010;14(2):209.
49. Su X, Johansen M, Looney MR, Brown EJ, Matthay MA. CD47 deficiency protects mice from lipopolysaccharide-induced acute lung injury and Escherichia coli pneumonia. J Immunol. 2008;180(10):6947–53.
50. Takeda Y, Suzuki M, Jin Y, Tachibana I. Preventive Role Of Tetraspanin CD9 in systemic inflammation of chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2015;53(6):751–60.
51. Ding J, Cui X, Liu Q. Emerging role of HMGB1 in lung diseases: friend or foe. J Cell Mol Med. 2017;21(6):1046–57.
52. Chen H, Qu J, Huang X, Kurundkar A, Zhu L, Yang N, et al. Mechanosensing by the α6-integrin confers an invasive fibroblast phenotype and mediates lung fibrosis. Nat Commun. 2016;7:12564.
53. Luzina IG, Todd NW, Nacu N, Lockatell V, Choi J, Hummers LK, et al. Regulation of pulmonary inflammation and fibrosis through expression of integrins αVβ3 and αVβ5 on pulmonary T lymphocytes. Arthritis Rheum. 2009;60(5):1530–9.
54. Rahman I, MacNee W. Role of transcription factors in inflammatory lung diseases. Thorax. 1998;53(7):601–12.
55. Fan J, Ye RD, Malik AB. Transcriptional mechanisms of acute lung injury. Am J Physiol Lung Cell Mol Physiol. 2001;281(5):L1037–50.
56. Cassel TN, Nord M. C/EBP transcription factors in the lung epithelium. Am J Physiol Lung Cell Mol Physiol. 2003;285(4):L773–81.
57. Qian L, Zhao Y, Guo L, Li S, Wu X. Activating transcription factor 3 (ATF3) protects against lipopolysaccharide-induced acute lung injury via inhibiting the expression of TL1A. J Cell Physiol. 2017;232(12):3727–34.
58. Bruno MD, Korfhagen TR, Liu C, Morrisey EE, Whitsett JA. GATA-6 activates transcription of surfactant protein A. J Biol Chem. 2000;275(2):1043–9.
59. Barnes PJ. Role of HDAC2 in the pathophysiology of COPD. Annu Rev Physiol. 2009;71:451–64.
60. Li L, Chong HC, Ng SY, Kwok KW, Teo Z, Tan EHP, et al. Angiopoietin-like 4 increases pulmonary tissue leakiness and damage during influenza pneumonia. Cell reports. 2015;10(5):654–63.
61. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020;181(5):1036–45.
62. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat Med. 2020;26(6):842–4.
63. Ahmad L, Mostowy S, Sancho-Shimizu V. Autophagy-virus interplay: from cell biology to human disease. Front Cell Dev Biol. 2018;6:155.
64. Yewdell JW, Hill AB. Viral interference with antigen presentation. Nat Immunol. 2002;3(11):1019–25.
65. Brandstadter JD, Yang Y. Natural killer cell responses to viral infection. Journal of innate immunity. 2011;3(3):274–9.
66. Olejnik J, Hume AJ, Mühlberger E. Toll-like receptor 4 in acute viral infection: Too much of a good thing. PLoS Pathog. 2018;14(12):e1007390.
67. Chen S, Sang N. Hypoxia-inducible factor-1: a critical player in the survival strategy of stressed cells. J Cell Biochem. 2016;117(2):267–78.
68. Alexiou K, Wilbring M, Matschke K, Dschietzig T. Relaxin protects rat lungs from ischemia-reperfusion injury via inducible NO synthase: role of ERK-1/2, PI3K, and forkhead transcription factor FKHRL1. PloS one. 2013;8(9):e75592-e.
69. Gurczynski SJ, Moore BB. IL-17 in the lung: the good, the bad, and the ugly. Am J Physiol Lung Cell Mol Physiol. 2018;314(1):L6–16.
70. Briguglio M, Bona A, Porta M, Dell'Osso B, Pregliasco FE, Banfi G. Disentangling the hypothesis of host dysosmia and SARS-CoV-2: the bait symptom that hides neglected neurophysiological routes. Front Physiol. 2020;11:671.
71. Atamas SP, Chapoval SP, Keegan AD. Cytokines in chronic respiratory diseases. F1000 Biol Rep. 2013;5:3.
72. Hou YJ, Okuda K, Edwards CE, Martinez DR, Asakura T, Dinnon KH, et al. SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. Cell. 2020;182(2):429-46.e14.
73. Ciulla MM. SARS-CoV-2 downregulation of ACE2 and pleiotropic effects of ACEIs/ARBs. Hypertension Research. 2020.
74. Cheng H, Wang Y, Wang G-Q. Organ-protective effect of angiotensin-converting enzyme 2 and its effect on the prognosis of COVID-19. J Med Virol. 2020;92(7):726–30.
75. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
76. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
77. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31(10):1674–6.
78. Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics. 2013;30(7):1003–5.
79. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160–5.
80. Zhao WM, Song SH, Chen ML, Zou D, Ma LN, Ma YK, et al. The 2019 novel coronavirus resource. Yi chuan Hereditas. 2020;42(2):212–21.
81. Corporation. M. Microsoft Excel. 2020. https://office.microsoft.com/excel.
82. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17(1):122.
83. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2012;41(D1):D991–5.
84. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom: Babraham Bioinformatics; 2010.
85. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
86. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
87. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 2010;38(12):e131.
88. Hubbard TJP, Aken BL, Beal K, Ballester B, Cáccamo M, Chen Y, et al. Ensembl 2007. Nucleic Acids Res. 2007;35(1):D610–7.
89. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41(10):e108.
90. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):106.
91. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25(3):415–6.
92. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance. 2017;22(13):30494.
93. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.

94. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
95. Perez-Llamas C, Lopez-Bigas N. Gitools: analysis and visualisation of genomic data using interactive heat-maps. PloS ONE. 2011;6:5.
96. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
97. Huang R, Grishagin I, Wang Y, Zhao T, Greene J, Obenauer JC, et al. The NCATS BioPlanet—an integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics. Front Pharmacol. 2019;10:445.
98. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
99. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2020;48(D1):D498-d503.
100. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B (Methodol). 1995;57(1):289–300.
101. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
102. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–13.

Publisher's Note

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