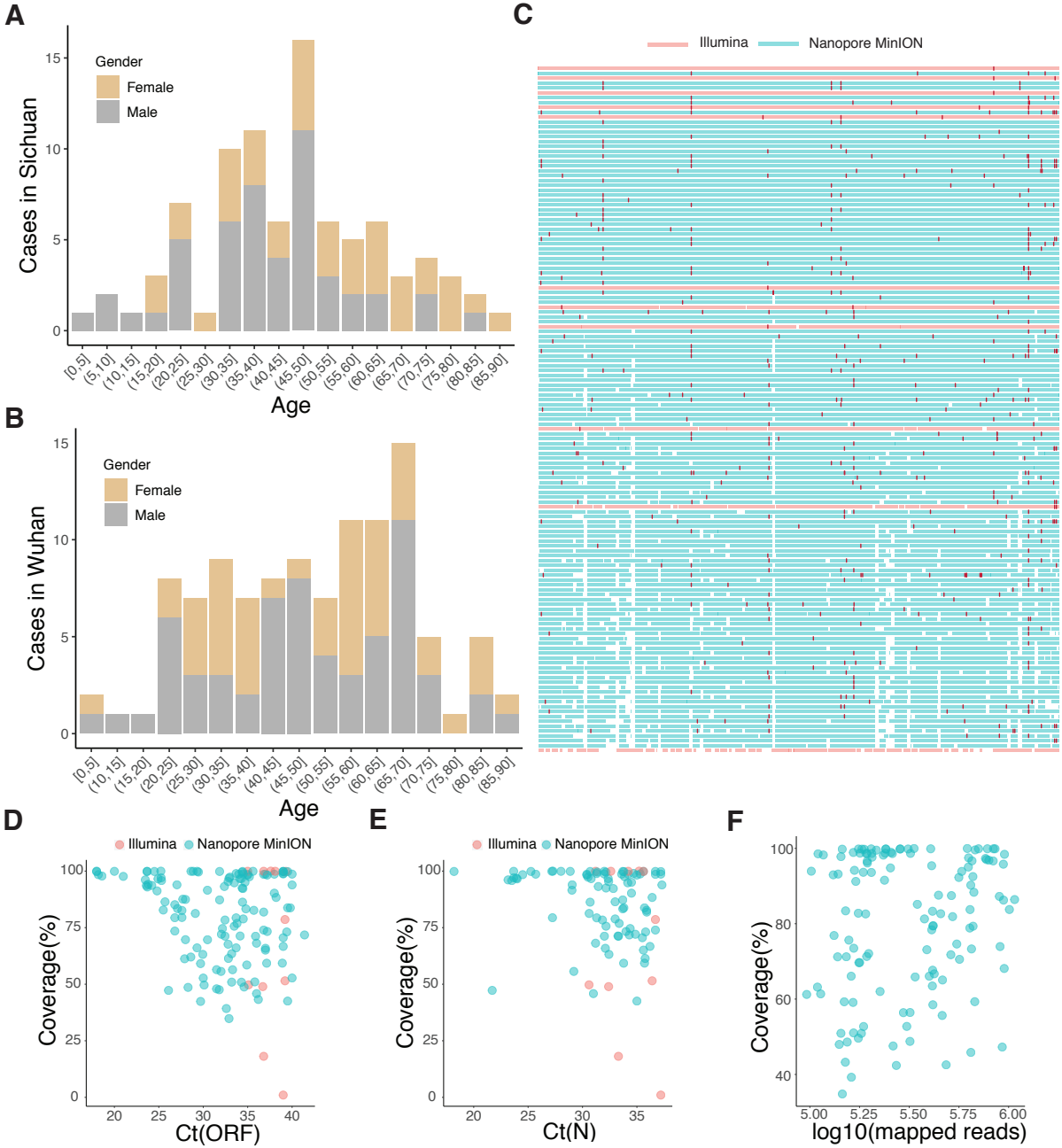
Cell Host & Microbe, Volume *29*

Supplemental information

Genomic monitoring of SARS-CoV-2 uncovers an Nsp1 deletion variant that modulates type I interferon response

Jing-wen Lin, Chao Tang, Han-cheng Wei, Baowen Du, Chuan Chen, Minjin Wang, Yongzhao Zhou, Ming-xia Yu, Lu Cheng, Suvi Kuivanen, Natacha S. Ogando, Lev Levanov, Yuancun Zhao, Chang-ling Li, Ran Zhou, Zhidan Li, Yiming Zhang, Ke Sun, Chengdi Wang, Li Chen, Xia Xiao, Xiuran Zheng, Sha-sha Chen, Zhen Zhou, Ruirui Yang, Dan Zhang, Mengying Xu, Junwei Song, Danrui Wang, Yupeng Li, ShiKun Lei, Wanqin Zeng, Qingxin Yang, Ping He, Yaoyao Zhang, Lifang Zhou, Ling Cao, Feng Luo, Huayi Liu, Liping Wang, Fei Ye, Ming Zhang, Mengjiao Li, Wei Fan, Xinqiong Li, Kaiju Li, Bowen Ke, Jiannan Xu, Huiping Yang, Shusen He, Ming Pan, Yichen Yan, Yi Zha, Lingyu Jiang, Changxiu Yu, Yingfen Liu, Zhiyong Xu, Qingfeng Li, Yongmei Jiang, Jiufeng Sun, Wei Hong, Hongping Wei, Guangwen Lu, Olli Vapalahti, Yunzi Luo, Yuquan Wei, Thomas Connor, Wenjie Tan, Eric J. Snijder, Teemu Smura, Weimin Li, Jia Geng, Binwu Ying, and Lu Chen


A
Cases in Sichuan
Gender
Female
Male
Age
B
Cases in Wuhan
Gender
Female
Male
Age
C
Illumina
Nanopore MinION
D
Illumina
Nanopore MinION
Coverage(%)
Ct(ORF)
E
Illumina
Nanopore MinION
Coverage(%)
Ct(N)
F
Coverage(%)
log10(mapped reads)

Supplementary Figures

Figure S1. Summary of two cohorts enrolled in this study and quality of generated genomes (related to **Figure 1**). Age and gender distribution of the COVID-19 cases in Sichuan Province (**A**) and Wuhan (**B**), China. Cases are classified according to their gender: grey indicates male and yellow indicates female. (**C**) Genome coverage map for the 141 genomes obtained from Sichuan cohort, ordered by genome coverage. SNPs are colored in dark red. Plots of SARS-CoV-2 genome coverage against Ct values of ORF (**D**) and N gene (**E**) in qPCR tests and log-transformed number of mapped reads (**F**). Genomes are colored according to the sequencing approach: blue indicates sequencing using a Nanopore MinION device, and pink indicates sequencing using Illumina NextSeq 500 platform.

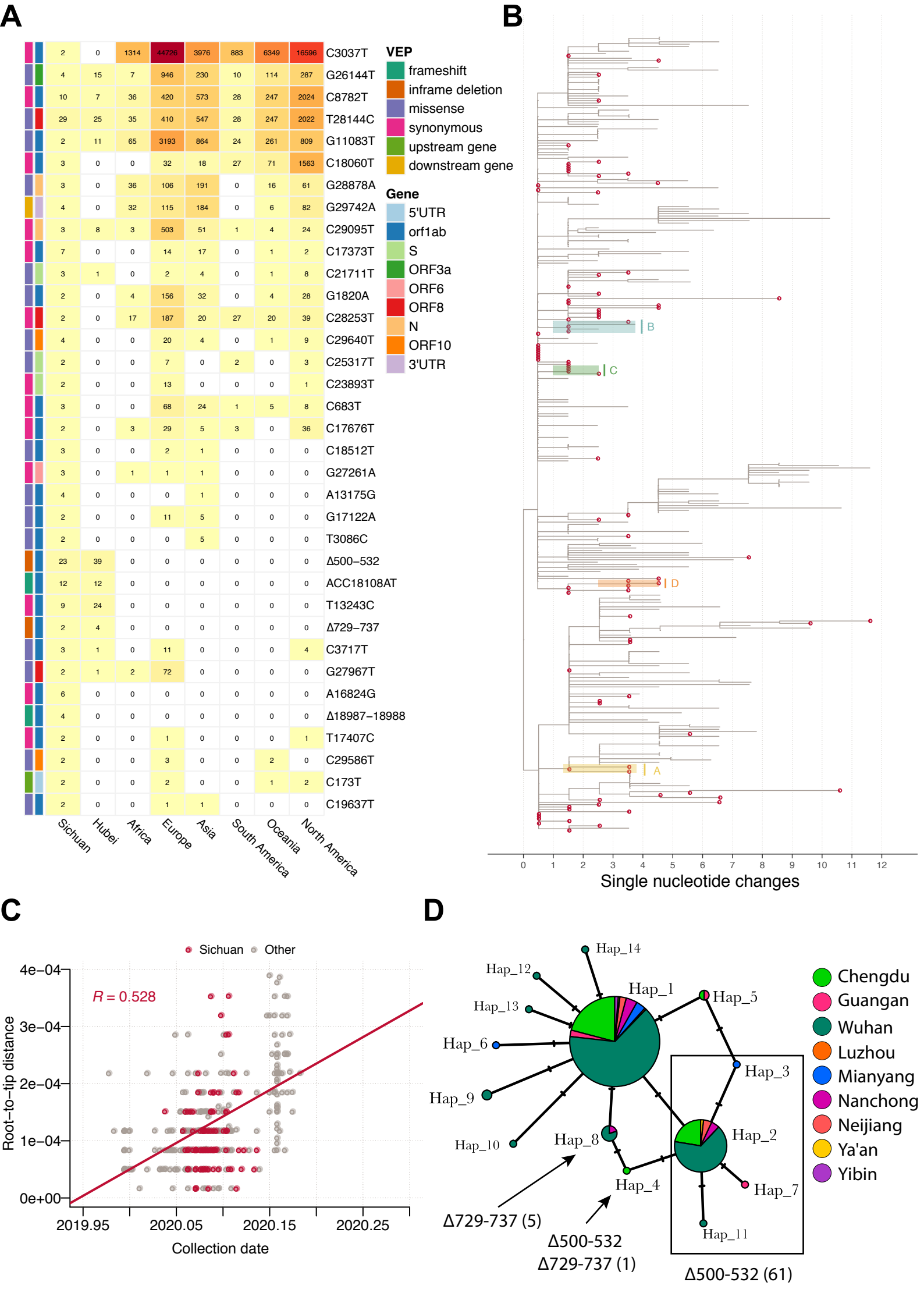

A
B
C
D
VEP
frameshift
inframe deletion
missense
synonymous
upstream gene
downstream gene
Gene
5'UTR
orf1ab
S
ORF3a
ORF6
ORF8
N
ORF10
3'UTR
C3037T
G26144T
C8782T
T28144C
G11083T
C18060T
G28878A
G29742A
C29095T
C17373T
C21711T
G1820A
C28253T
C29640T
C25317T
C23893T
C683T
C17676T
C18512T
G27261A
A13175G
G17122A
T3086C
Δ500–532
ACC18108AT
T13243C
Δ729–737
C3717T
G27967T
A16824G
Δ18987–18988
T17407C
C29586T
C173T
C19637T
Sichuan
Hubei
Africa
Europe
Asia
South America
Oceania
North America
Single nucleotide changes
Sichuan
Other
R = 0.528
Root–to–tip distance
Collection date
Hap_14
Hap_12
Hap_1
Hap_5
Hap_13
Hap_6
Hap_3
Hap_9
Hap_10
Hap_8
Hap_2
Hap_4
Hap_7
Hap_11
Δ729-737 (5)
Δ500-532
Δ729-737 (1)
Δ500-532 (61)
Chengdu
Guangan
Wuhan
Luzhou
Mianyang
Nanchong
Neijiang
Ya'an
Yibin

Figure S2. Identified recurrent genetic variants, phylogenetic and haplotype analysis (related to **Figure 2**). (**A**) Number of samples with the 35 recurrent genetic variants in China (Sichuan, Hubei, Guangdong or Hong Kong) or other countries. The effect of the variants was predicted by VEP (Variant Effect Predictor). (**B**) Estimated maximum likelihood phylogenetic tree using SARS-CoV-2 genome sequences from Sichuan (red circles), other provinces of China and other countries (grey branches). The x-axis shows the number of nucleotide changes from the inferred root sequence. The position of clusters A-D were labelled and shaded with colors. An interactive website presenting the lineage, location and country name can be accessed at http://covid19.chenlulab.com/dist/tree.html. (**C**) The correlation between the collection date and the root-to-tip distance in the maximum likelihood tree. Each dot represents a SARS-COV-2 genome. Red dots are virus genomes from Sichuan. The y-axis is the root-to-tip distance extracted from the phylogenetic trees generated by PhyML, with the root placed by TempEst. The x-axis is the actual sampling dates of the samples. The red line shows the association between these two and the Pearson correlation coefficient is 0.528. (**D**) The Nsp1 haplotype network of SARS-CoV-2 viruses from Wuhan and 8 cities in Sichuan. The size of dots is proportional to the number of haplotype obtained. Haplotype ID is shown above each dot and black line connecting haplotypes with a vertical line representing one mutational change. Two deletions (Δ500-532, Δ729-737) and their number of cases are highlighted. All haplotypes inside the rectangle contain the deletion Δ500-532.

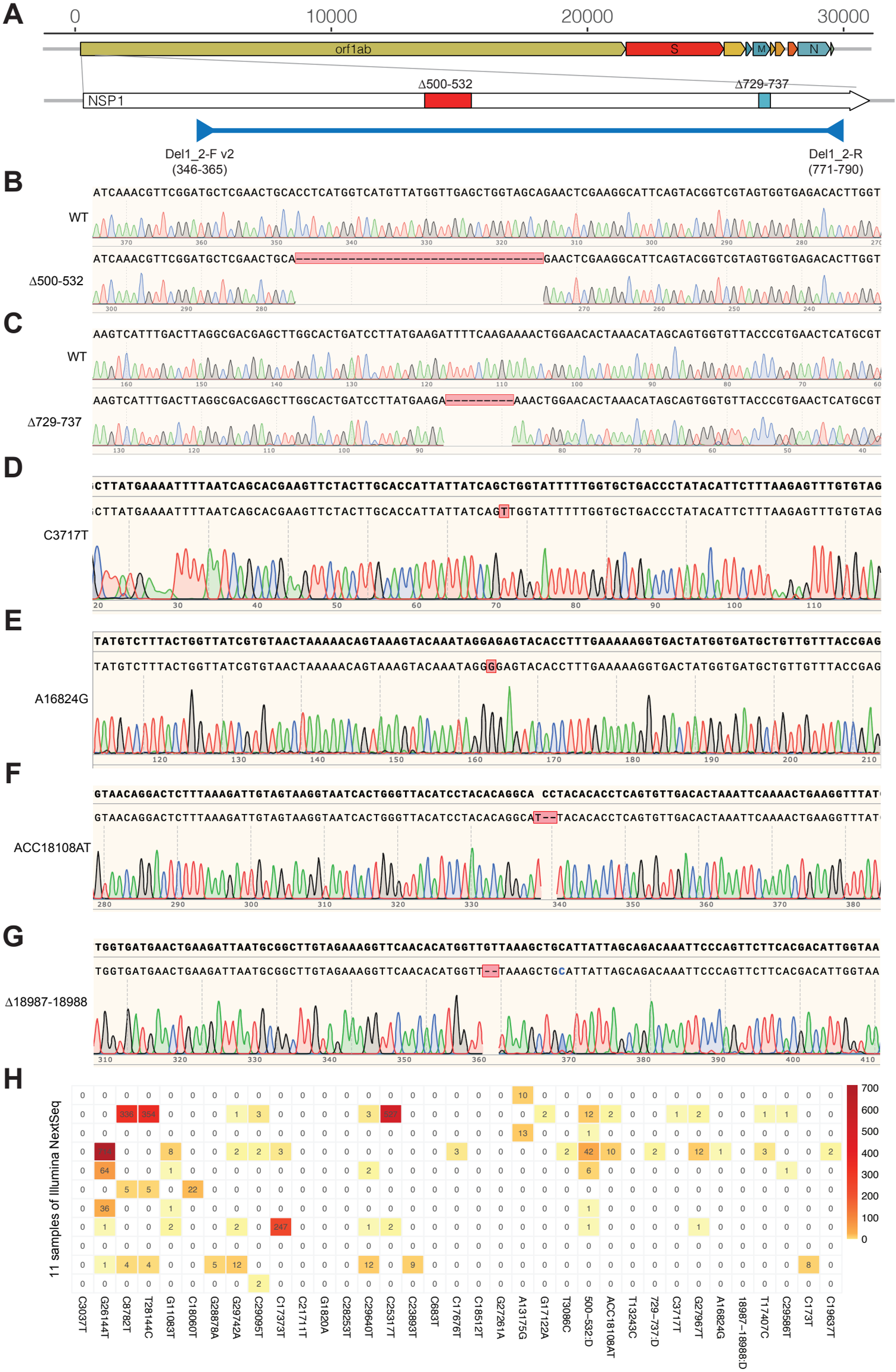
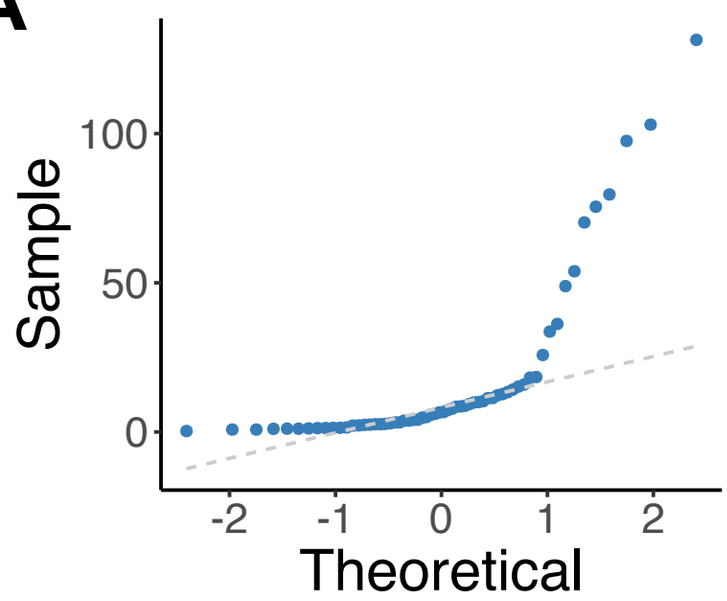
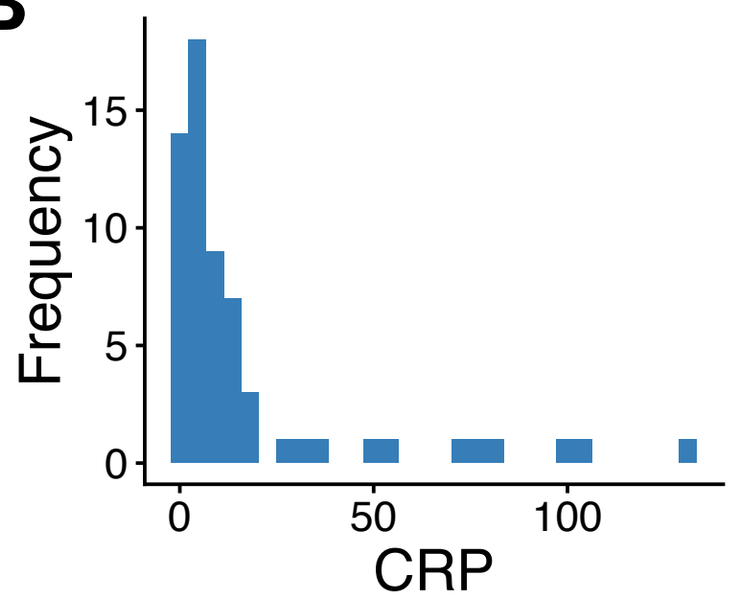

A
0
10000
20000
30000
orf1ab
S
M
N
NSP1
Δ500-532
Δ729-737
Del1_2-F v2
(346-365)
Del1_2-R
(771-790)
B
WT
Δ500-532
C
WT
Δ729-737
D
C3717T
E
A16824G
F
ACC18108AT
G
Δ18987-18988
H
11 samples of Illumina NextSeq
C3037T
G26144T
C8782T
T28144C
G11083T
C18060T
G28878A
G29742A
C29095T
C17373T
C21711T
G1820A
C28253T
C29640T
C25317T
C23893T
C683T
C17676T
C18512T
G27261A
A13175G
G17122A
T3086C
500–532:D
ACC18108AT
T13243C
729–737:D
C3717T
G27967T
A16824G
18987–18988:D
T17407C
C29586T
C173T
C19637T
700
600
500
400
300
200
100
0

Figure S3. Validation of deletions and SNPs (related to **Figure 2**). (**A**) Location of Nsp1 coding region in the SARS-CoV-2 genome, locations of the two deletions (Δ500-532 and Δ729-737) in Nsp1 and primers designed for PCR. Sanger sequencing confirms the two deletions Δ500-532 (**B**) and Δ729-737 (**C**). Four variants, including C3717T (**D**), A16824G (**E**), ACC18108AT (**F**) and Δ18987-18988 (**G**) were also validated by Sanger sequencing. (**H**) Eleven samples that were sequenced using Illumina NextSeq platform were used to cross-validate the 35 recurrent variants. The numbers indicate the counts from each sample that support the corresponding SNP or deletion, from which 26 out of the 35 variants were also identified with Illumina sequencing.

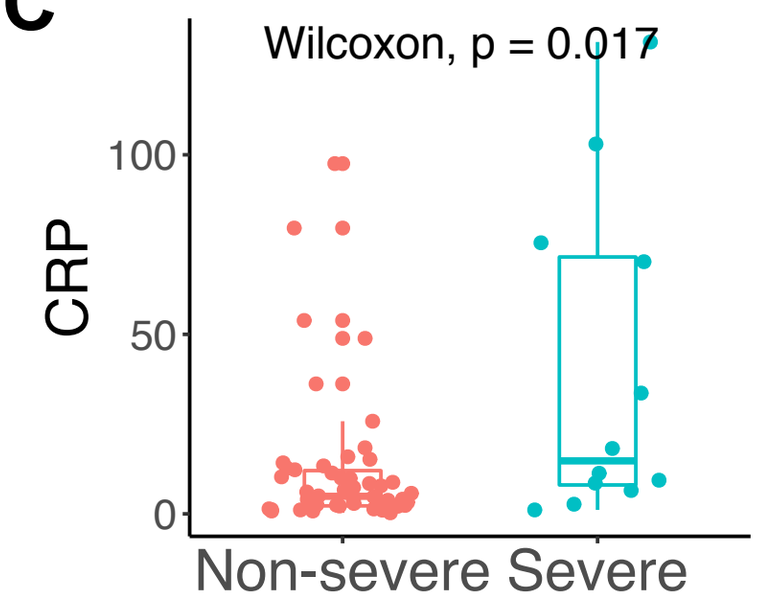
A


Sample
100
50
0
-2
-1
0
1
2
Theoretical

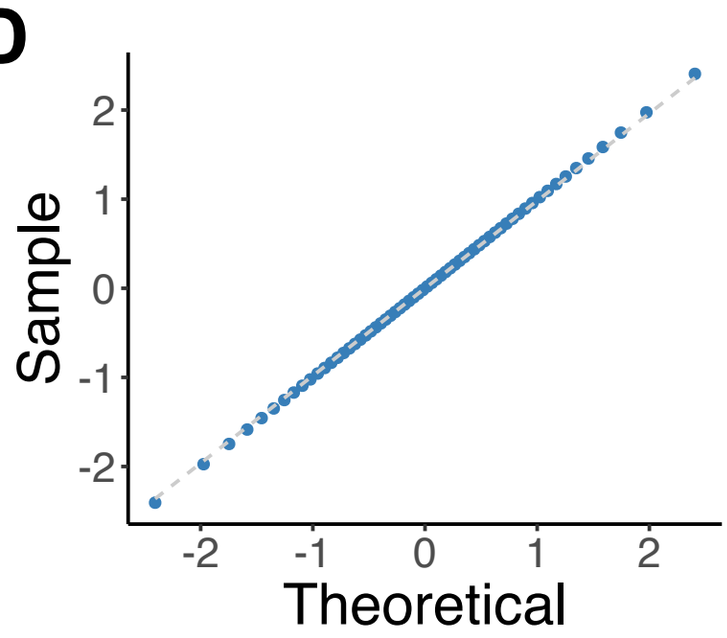
B


Frequency
15
10
5
0
0
50
100
CRP

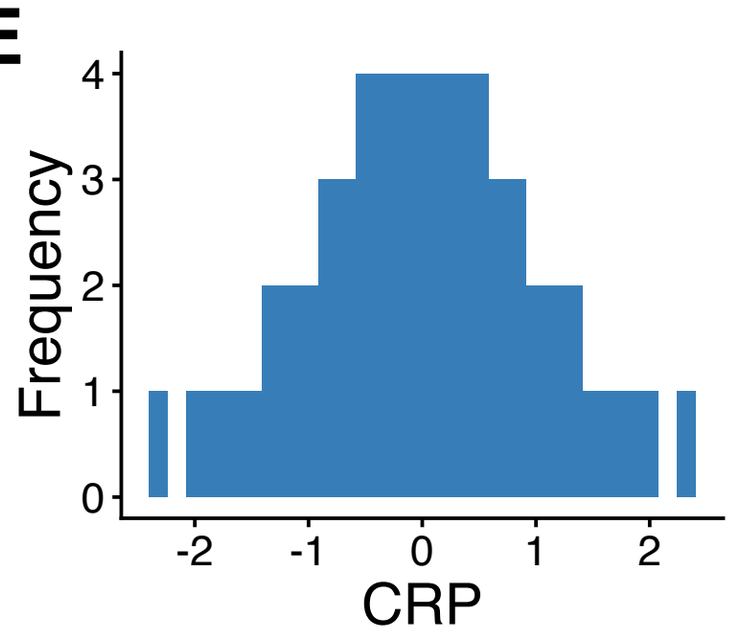
C


Wilcoxon, p = 0.017
CRP
100
50
0
Non-severe
Severe

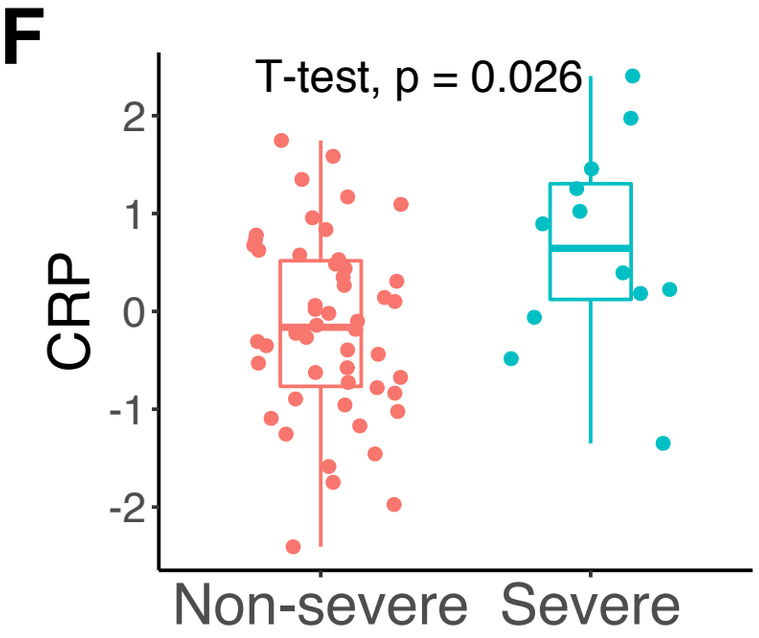
D


Sample
2
1
0
-1
-2
-2
-1
0
1
2
Theoretical

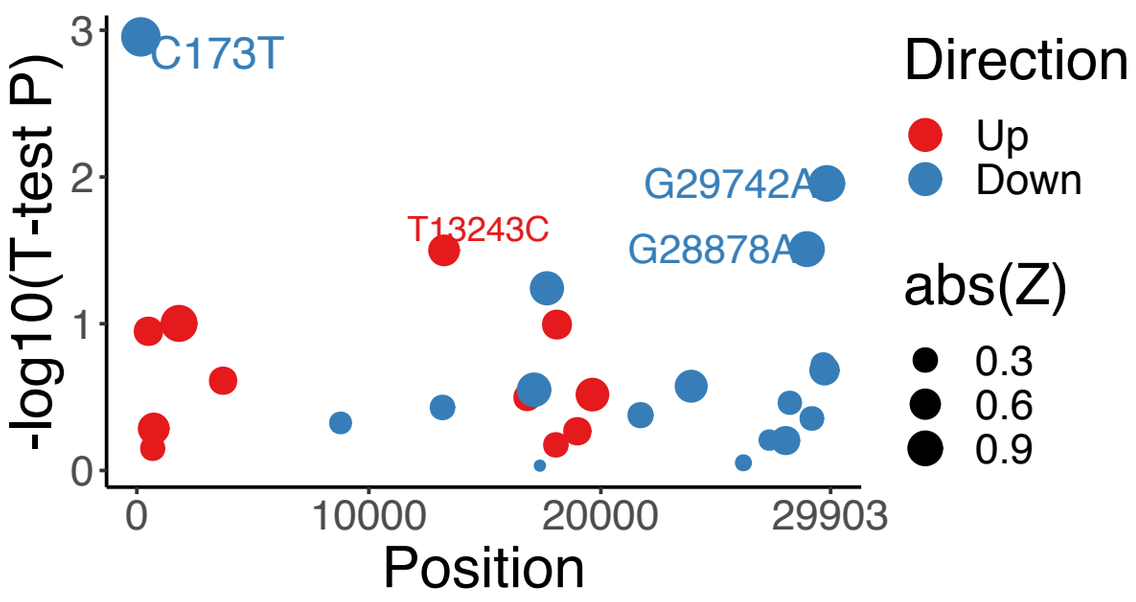
E


Frequency
4
3
2
1
0
-2
-1
0
1
2
CRP

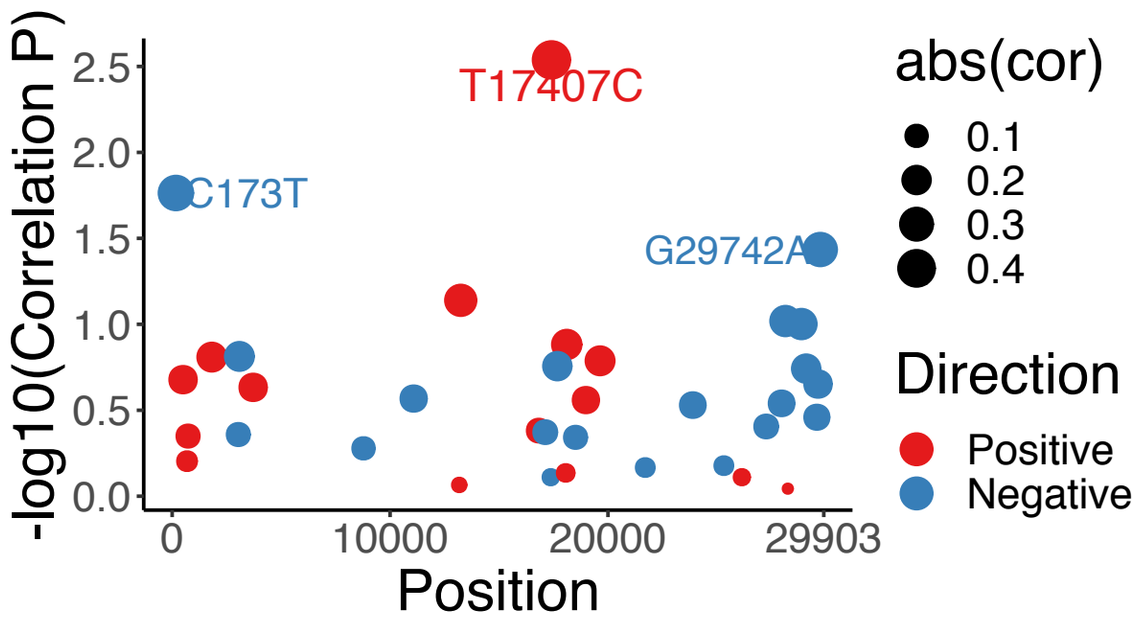
F


T-test, p = 0.026
CRP
2
1
0
-1
-2
Non-severe
Severe

G


-log10(T-test P)
3
2
1
0
C173T
T13243C
G29742A
G28878A
0
10000
20000
29903
Position
Direction
Up
Down
abs(Z)
0.3
0.6
0.9

H


-log10(Correlation P)
2.5
2.0
1.5
1.0
0.5
0.0
T17407C
C173T
G29742A
0
10000
20000
29903
Position
abs(cor)
0.1
0.2
0.3
0.4
Direction
Positive
Negative

I

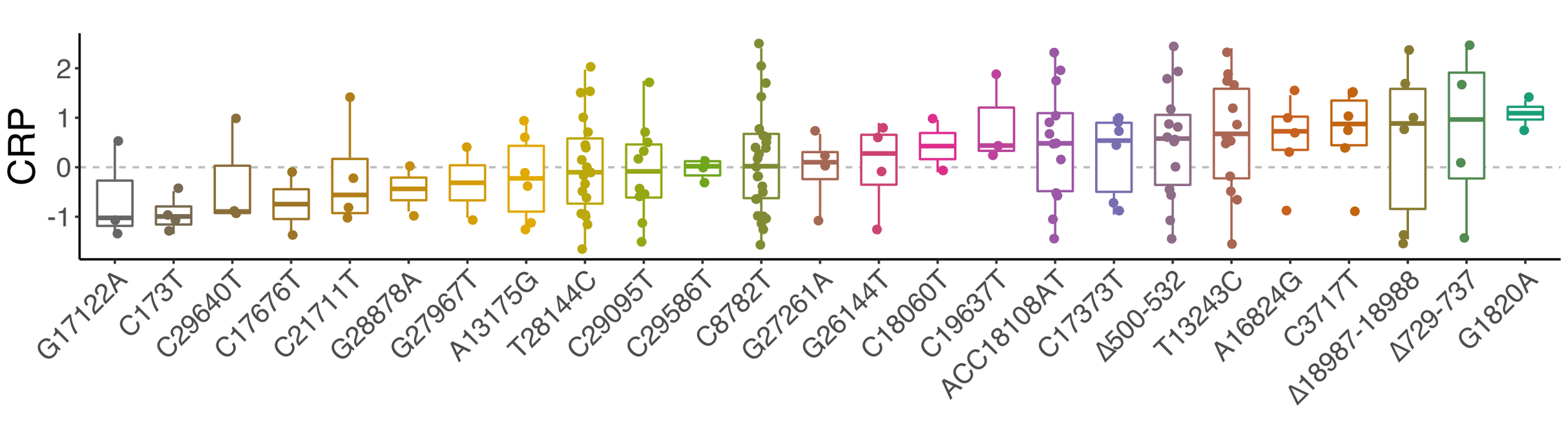

CRP
2
1
0
-1
G17122A
C173T
C29640T
C17676T
C21711T
G28878A
G27967T
A13175G
T28144C
C29095T
C29586T
C8782T
G27261A
G26144T
C18060T
C19637T
ACC18108AT
C17373T
Δ500-532
T13243C
A16824G
C3717T
Δ18987-18988
Δ729-737
G1820A

Figure S4. Normalization and association study of C-reactive protein (CRP) with 35 genetic variants (related to **Figure 3**). Quantile-quantile (QQ) plot and distribution plot before (**A** and **B**) and after normalization (**D** and **E**). An inverse normal transformation was used. Boxplots of CRP in non-severe and severe COVID-19 patients before (**C**) and after (**F**) normalization. P values of the Wilcoxon test or T-test were shown. (**G**) Manhattan plot of 35 genetic variants that were associated with CRP using T-test between severe and non-severe groups. Genome position and –log10 of P values were shown. Variants with P value below 0.05 were labeled. Red and blue indicate higher and lower values in severe group, respectively. The size of the dots is proportional to the absolute Z score number. (**H**) Manhattan plot of 35 genetic variants that were associated with CRP using Pearson correlation. Red and blue indicate positive and negative correlation, respectively. The size of the dots represents the absolute number of correlation coefficient (cor). (**I**) Boxplot of the sorted Z score values of CRP after normalization of 35 recurrent genetic variants.

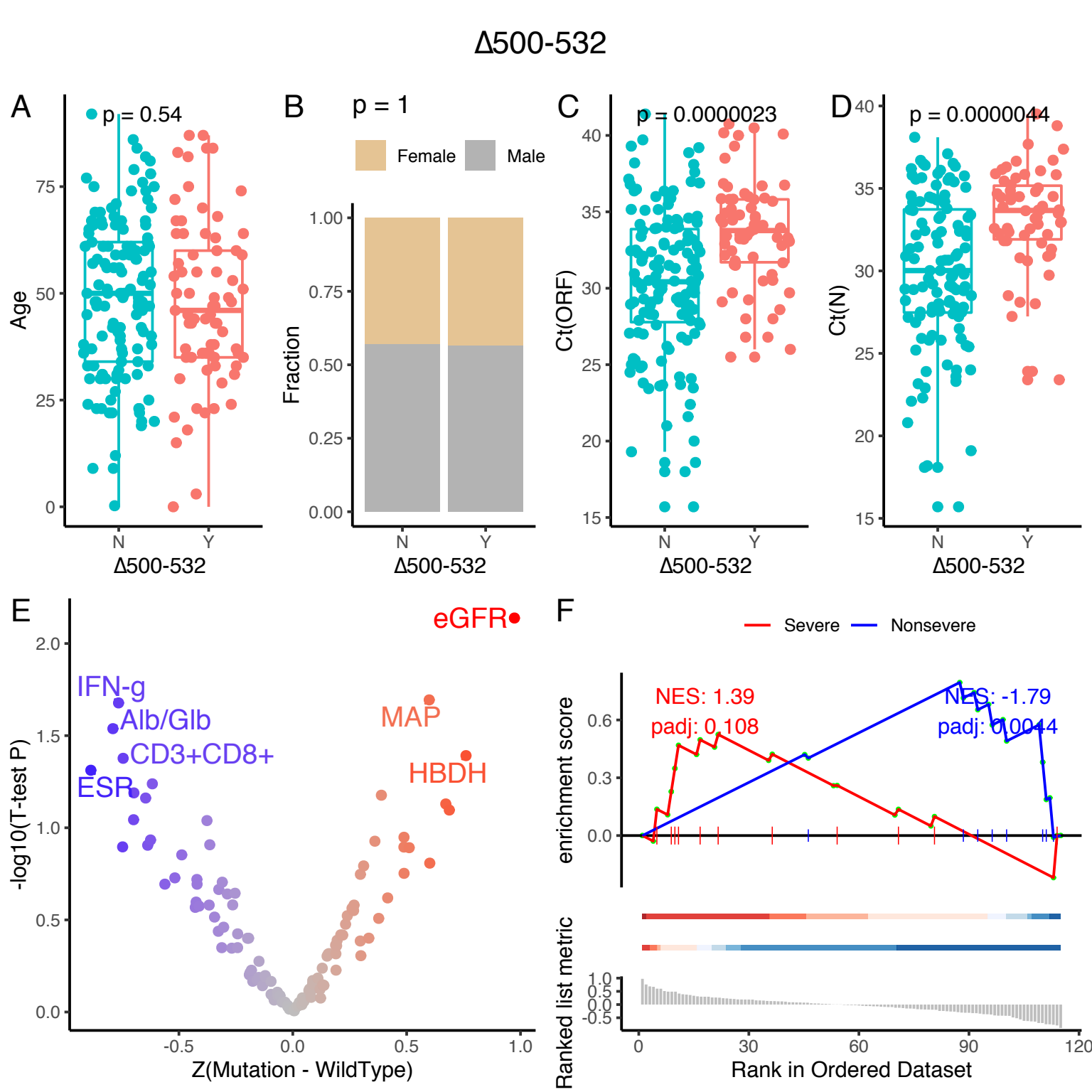

Δ500-532
A
p = 0.54
Age
N
Y
Δ500-532
B
p = 1
Female
Male
Fraction
N
Y
Δ500-532
C
p = 0.0000023
Ct(ORF)
N
Y
Δ500-532
D
p = 0.0000044
Ct(N)
N
Y
Δ500-532
E
eGFR
IFN-g
Alb/Glb
CD3+CD8+
ESR
MAP
HBDH
-log10(T-test P)
Z(Mutation - WildType)
F
Severe
Nonsevere
NES: 1.39
padj: 0.108
NES: -1.79
padj: 0.0044
enrichment score
Ranked list metric
Rank in Ordered Dataset

Figure S5. Comparison between two groups of patients with or without Δ500-532 mutant virus (related to **Figure 3**). Age (**A**), gender (**B**), Ct values of ORF (**C**) and N gene (**D**) in qPCR tests were compared between groups of patients with (Y) or without (N) Δ500-532 mutant virus. P values for age and Ct values (Wilcox-tests), and gender (chi-square test) were shown. (**E**) The volcano plot shows clinical phenotypes that were significantly different between the two groups of patients with or without Δ500-532. The y-axis is the –log10 p values of phenotypes the traits with p values less than 0.05 were shown. The x-axis shows the difference of Z score between cases with mutant or wildtype virus. See also **Table S8**. Alb/Glb, the albumin/globulin ratio; $CD3^+CD8^+$, $CD3^+CD8^+$ cell count; ESR, erythrocyte sedimentation rate; eGFR, estimated glomerular filtration rate; MAP, mean arterial pressure; HBDH, α-Hydroxybutyrate dehydrogenase. (**F**) Ranked enrichment analysis of severe and non-severe clinical phenotypes for Δ500-532. NES, normalized enrichment score. P values were adjusted using permutation.

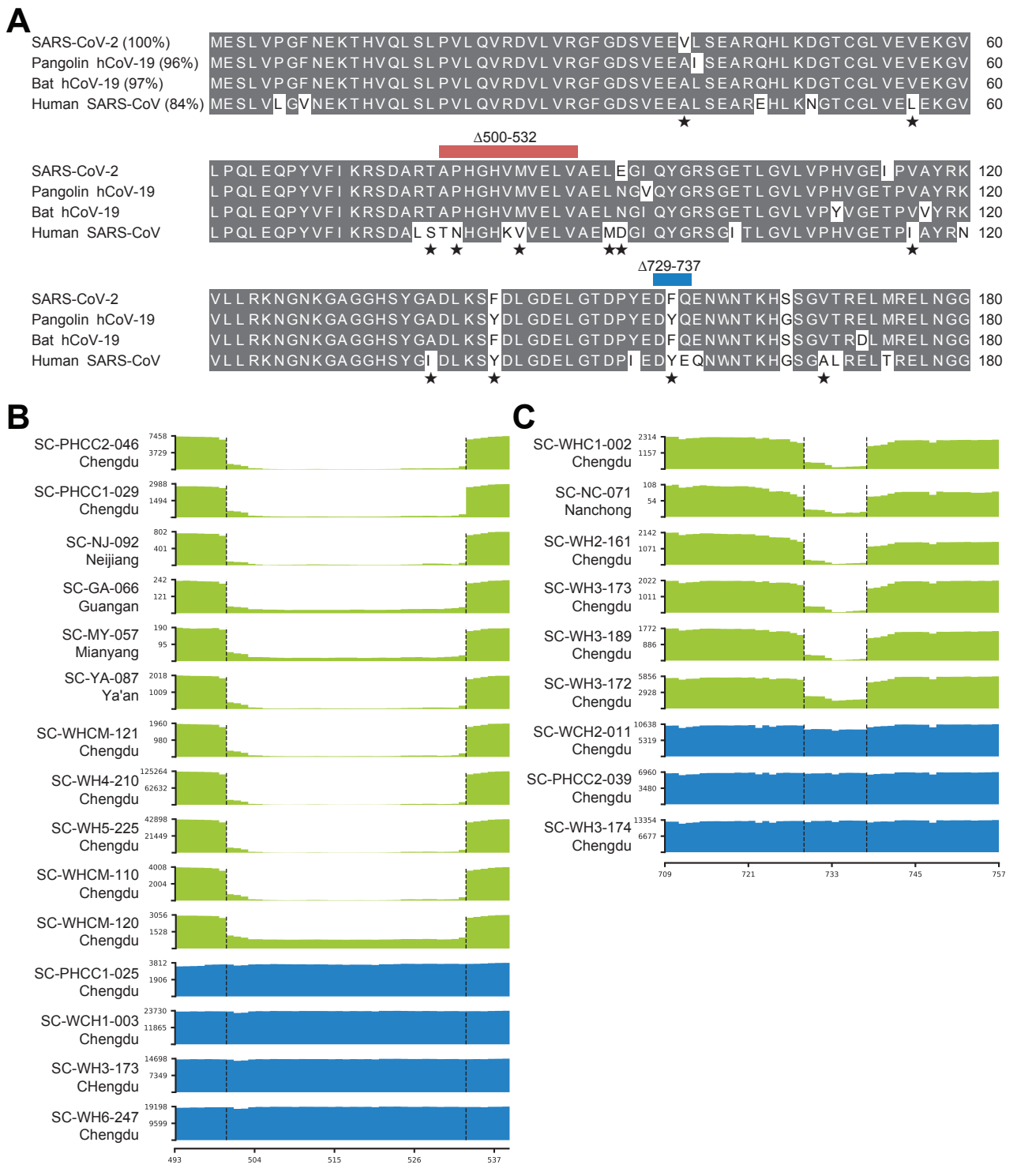

A
SARS-CoV-2 (100%)
Pangolin hCoV-19 (96%)
Bat hCoV-19 (97%)
Human SARS-CoV (84%)
60
Δ500-532
SARS-CoV-2
Pangolin hCoV-19
Bat hCoV-19
Human SARS-CoV
120
Δ729-737
180
B
SC-PHCC2-046 Chengdu
SC-PHCC1-029 Chengdu
SC-NJ-092 Neijiang
SC-GA-066 Guangan
SC-MY-057 Mianyang
SC-YA-087 Ya'an
SC-WHCM-121 Chengdu
SC-WH4-210 Chengdu
SC-WH5-225 Chengdu
SC-WHCM-110 Chengdu
SC-WHCM-120 Chengdu
SC-PHCC1-025 Chengdu
SC-WCH1-003 Chengdu
SC-WH3-173 CHengdu
SC-WH6-247 Chengdu
493
504
515
526
537
C
SC-WHC1-002 Chengdu
SC-NC-071 Nanchong
SC-WH2-161 Chengdu
SC-WH3-173 Chengdu
SC-WH3-189 Chengdu
SC-WH3-172 Chengdu
SC-WCH2-011 Chengdu
SC-PHCC2-039 Chengdu
SC-WH3-174 Chengdu
709
721
733
745
757

Figure S6. Sequence alignment of Nsp1 proteins and genome coverage of two deletions in Nsp1 (related to **Figure 4** and **5**). (**A**) Alignment of Nsp1 amino acid sequences. Stars indicate amino acid substitutions that are of similar types. The percentages of amino acid identical to SARS-CoV-2 are shown in the brackets and the locations of the two deletions identified in this study are indicated. Sashimi plots of representative samples with (**B**) **Δ**500-532 or (**C**) **Δ**729-737 mutation (green), compared to samples without the corresponding deletion (blue).

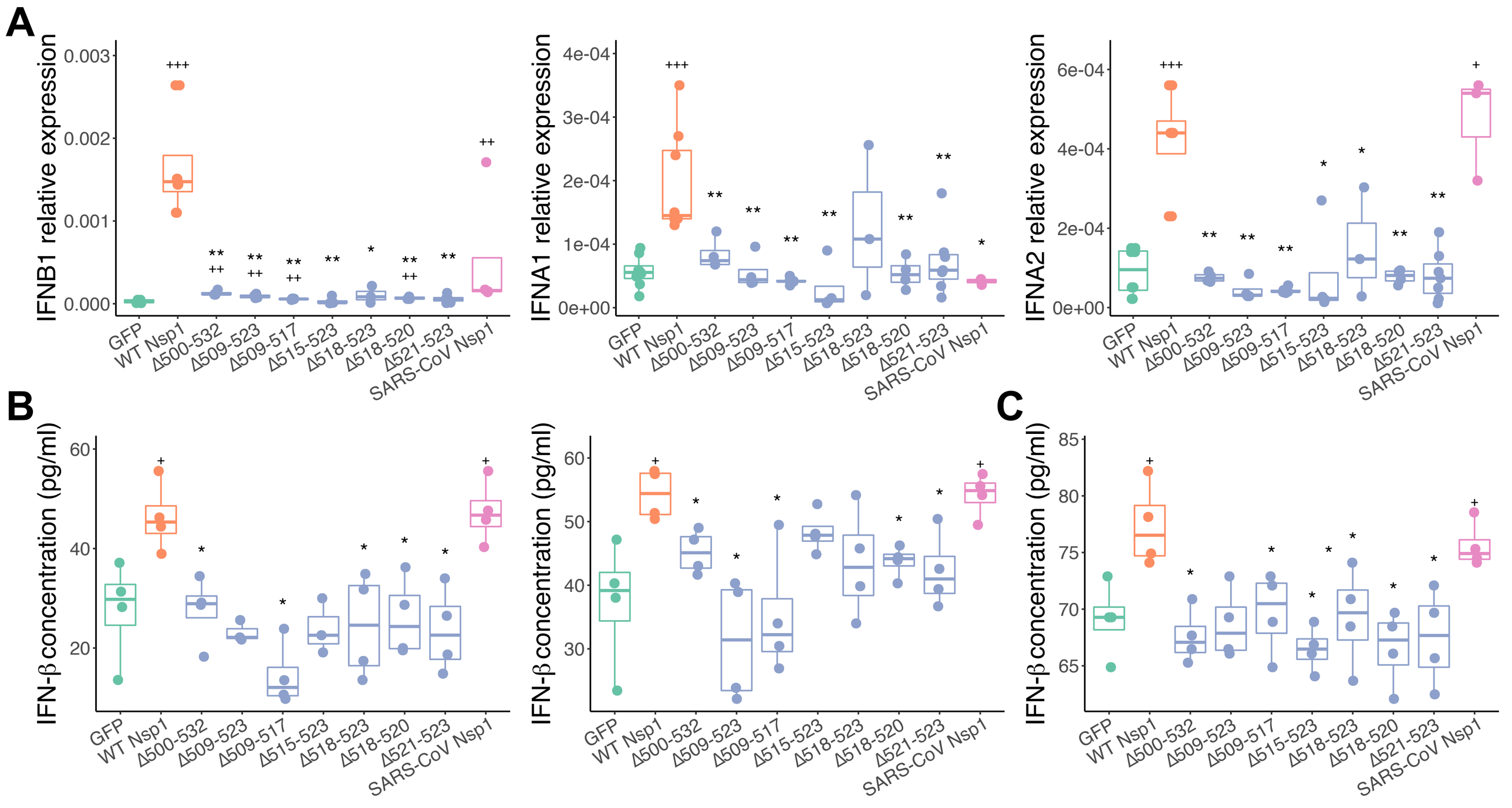

A
IFNB1 relative expression
0.003
0.002
0.001
0.000
IFNA1 relative expression
4e-04
3e-04
2e-04
1e-04
0e+00
IFNA2 relative expression
6e-04
4e-04
2e-04
0e+00
GFP
WT Nsp1
Δ500-532
Δ509-523
Δ509-517
Δ515-523
Δ518-523
Δ518-520
Δ521-523
SARS-CoV Nsp1
B
IFN-β concentration (pg/ml)
60
40
20
C
85
80
75
70
65

Figure S7. Nsp1 mutants down-regulate IFN-I response (related to **Figure 6**).

(**A**) Relative mRNA expression level of IFNB1, IFNA1 and IFNA2 to GAPDH in SARS-Cov-2 wildtype (WT) or mutant Nsp1-expressing HEK293T cells. (**B**) Concentration of IFN-β in the supernatant (left) or cell lysate (right) of Nsp1-expressing HEK293T cells. (**C**) Concentration of IFN-β in cell lysate of Nsp1-expressing A549 cells. GFP and SARS-CoV Nsp1 were used as negative and positive controls. + indicates statistical significance compared to GFP controls and * indicates significance compared to WT SARS-CoV-2 Nsp1, Mann-Whitney U test was used.