nature

https://doi.org/10.1038/s41586-020-2012-7

Accelerated Article Preview

A pneumonia outbreak associated with a new coronavirus of probable bat origin

Received: 20 January 2020

Accepted: 29 January 2020

Accelerated Article Preview
Published online 3 February 2020

Cite this article as: Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* https://doi.org/10.1038/s41586-020-2012-7 (2020).

Open access

Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao & Zheng-Li Shi

A pneumonia outbreak associated with a new coronavirus of probable bat origin

https://doi.org/10.1038/s41586-020-2012-7

Received: 20 January 2020

Accepted: 29 January 2020

Published online: 3 February 2020

Open access

Peng Zhou[1,5], Xing-Lou Yang[1,5], Xian-Guang Wang[2,5], Ben Hu[1], Lei Zhang[1], Wei Zhang[1], Hao-Rui Si[1,3], Yan Zhu[1], Bei Li[1], Chao-Lin Huang[2], Hui-Dong Chen[2], Jing Chen[1,3], Yun Luo[1,3], Hua Guo[1,3], Ren-Di Jiang[1,3], Mei-Qin Liu[1,3], Ying Chen[1,3], Xu-Rui Shen[1,3], Xi Wang[1,3], Xiao-Shuang Zheng[1,3], Kai Zhao[1,3], Quan-Jiao Chen[1], Fei Deng[1], Lin-Lin Liu[4], Bing Yan[1], Fa-Xian Zhan[4], Yan-Yi Wang[1], Geng-Fu Xiao[1] & Zheng-Li Shi[1]*

Since the SARS outbreak 18 years ago, a large number of severe acute respiratory syndrome-related coronaviruses (SARSr-CoV) have been discovered in their natural reservoir host, bats[1–4]. Previous studies indicated that some of those bat SARSr-CoVs have the potential to infect humans[5–7]. Here we report the identification and characterization of a novel coronavirus (2019-nCoV) which caused an epidemic of acute respiratory syndrome in humans in Wuhan, China. The epidemic, which started from 12 December 2019, has caused 2,050 laboratory-confirmed infections with 56 fatal cases by 26 January 2020. Full-length genome sequences were obtained from five patients at the early stage of the outbreak. They are almost identical to each other and share 79.5% sequence identify to SARS-CoV. Furthermore, it was found that 2019-nCoV is 96% identical at the whole-genome level to a bat coronavirus. The pairwise protein sequence analysis of seven conserved non-structural proteins show that this virus belongs to the species of SARSr-CoV. The 2019-nCoV virus was then isolated from the bronchoalveolar lavage fluid of a critically ill patient, which can be neutralized by sera from several patients. Importantly, we have confirmed that this novel CoV uses the same cell entry receptor, ACE2, as SARS-CoV.

Coronavirus has caused two large-scale pandemic in the last two decades, SARS and MERS (Middle East respiratory syndrome)[8,9]. It was generally believed that SARSr-CoV, mainly found in bats, might cause future disease outbreak[10,11]. Here we report on a series of unidentified pneumonia disease outbreaks in Wuhan, Hubei province, central China. Started from a local seafood market, the outbreak has grown substantial to infect 2050 people in China with 56 deaths and to infect 35 people in 11 other countries up to January 26, 2020[12]. Typical clinical symptoms of these patients are fever, dry cough, dyspnea, headache, and pneumonia. Disease onset may result in progressive respiratory failure due to alveolar damage (as observed by transverse chest CT images) and even death. The disease was determined as viral induced pneumonia by clinicians according to clinical symptoms and other criteria including body temperature rising, lymphocytes and white blood cells decreasing (sometimes normal for the later), new pulmonary infiltrates on chest radiography, and no obvious improvement upon three days antibiotics treatment. It appears most of the early cases had contact history with the original seafood market, but the disease progressed to human-to-human transmission now.

Samples from seven patients with severe pneumonia (six are seafood market sellers or delivers), who were enrolled in intensive unit cares at the beginning of the outbreak, were sent to WIV laboratory for pathogen diagnosis (Extended Data Table 1). As a CoV lab, we first used pan-CoV PCR primers to test these samples[13], considering the outbreak happened in winter and in a market, same environment as SARS. We found five PCR positive. A sample (WIV04) collected from bronchoalveolar lavage fluid (BALF) was analysed by metagenomics analysis using next-generation sequencing (NGS) to identify potential etiological agents. Of the 10,038,758 total reads, or 1582 total reads obtained after human genome filtering, 1378 (87.1%) matched sequences of SARSr-CoV (Fig. 1a). By *de novo* assembly and targeted PCR, we obtained a 29,891-bp CoV genome that shared 79.5% sequence identity to SARS-CoV BJ01 (GenBank accession number AY278488.2). High genome coverage was obtained by remapping the total reads to this genome (Extended Data Figure 1). This sequence has been submitted to GISAID (accession no. EPI_ISL_402124). Following the name by WHO, we tentatively call it novel coronavirus 2019 (2019-nCoV). Four more full-length genome sequences of 2019-nCoV (WIV02, WIV05, WIV06, and WIV07) (GISAID accession nos. EPI_ISL_402127-402130) that were above 99.9% identical to each other were subsequently obtained from other four patients using NGS and PCR (Extended Data Table 2).

The virus genome consists of six major open reading frames (ORFs) common to coronaviruses and a number of other accessory genes (Fig. 1b). Further analysis indicates that some of the 2019-nCoV genes shared less than 80% nt sequence identity to SARS-CoV. However, the seven conserved replicase domains in ORF1ab that were used for

[1]CAS Key Laboratory of Special Pathogens, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, People's Republic of China. [2]Wuhan Jinyintan Hospital, Wuhan, China. [3]University of Chinese Academy of Sciences, Beijing, People's Republic of China. [4]Hubei Provincial Center for Disease Control and Prevention, Wuhan, People's Republic of China. [5]These authors contributed equally: Peng Zhou, Xing-Lou Yang, Xian-Guang Wang. *e-mail: zlshi@wh.iov.cn

CoV species classification, are 94.6% aa sequence identical between 2019-nCoV and SARS-CoV, implying the two belong to same species (Extended Data Table 3).

We then found a short RdRp region from a bat coronavirus termed BatCoV RaTG13 which we previously detected in *Rhinolophus affinis* from Yunnan Province showed high sequence identity to 2019-nCoV. We did full-length sequencing to this RNA sample (GISAID accession no.EPI_ISL_402131). Simplot analysis showed that 2019-nCoV was highly similar throughout the genome to RaTG13 (Fig. 1c), with 96.2% overall genome sequence identity. Using the aligned genome sequences of 2019-nCoV, RaTG13, SARS-CoV and previously reported bat SARSr-CoVs, no evidence for recombination events was detected in the genome of 2019-nCoV. The phylogenetic analysis of full-length genome, RNA-dependent RNA polymerase (RdRp) gene and S gene sequences all showed that RaTG13 is the closest relative of the 2019-nCoV and form a distinct lineage from other SARSr-CoVs (Fig. 1d and Extended Data Figure 2). The receptor binding protein spike (S) gene was highly divergent to other CoVs (Extended Data Figure 2), with less than 75% nt sequence identity to all previously described SARSr-CoVs except a 93.1% nt identity to RaTG13 (Extended Data Table 3). The S genes of 2019-nCoV and RaTG13 S gene are longer than other SARSr-CoVs. The major differences in 2019-nCoV are the three short insertions in the N-terminal domain, and four out of five key residues changes in the receptor-binding motif, in comparison with SARS-CoV (Extended Data Figure 3). Whether the insertions in N-terminal domain of 2019-nCoV confers a sialic acid binding activity like MERS-CoV needs to be further studied. The close phylogenetic relationship to RaTG13 provides evidence for a bat origin of 2019-nCoV.

We rapidly developed a qPCR detection based on the receptor-binding domain of spike gene, the most variable region among genome (Fig. 1c). Our data show the primers could differentiate 2019-nCoV with all other human coronaviruses including bat SARSr-CoV WIV1, which is 95% identity to SARS-CoV (Extended Data Figure 4a and 4b). From the seven patients, we found 2019-nCoV positive in six BALF and five oral swab samples during the first sampling by qPCR and conventional PCR. However, we can no longer find viral positive in oral swabs, anal swabs, and blood from these patients during the second sampling (Fig. 2a). We have to point out that other qPCR targets including RdRp or E gene may be suggested for routine detection. Based on these findings, we presume that the disease should be transmitted through airway, yet we can't rule out other possibilities if the investigation extended to include more patients.

For serological detection of 2019-nCoV, we used previously developed bat SARSr-CoV Rp3 nucleocapsid protein (NP) as antigen in IgG and IgM ELISA test, which shared 92% amino acid identity to 2019-nCoV NP (Extended Data Figure 5) and showed no cross-reactivity against other human coronaviruses except SARSr-CoV[7]. As a research lab, we were only able to get five serum samples from the seven viral infected patients. We monitored viral antibody levels in one patient (ICU-06) at seven, eight, nine, and eighteen days after disease onset (Extended Data Table 2). A clear trend of IgG and IgM titre (decreased at the last day) increase was observed (Fig. 2b). For a second investigation, we tested viral antibody for five of the seven viral positive patients around twenty days after disease onset (Extended Data Table 1 and 2). All patient samples, but not samples from healthy people, showed strong viral IgG positive (Fig. 2b). We also found three IgM positive, indicating acute infection.

We then successfully isolated the virus (named 2019-nCoV BetaCoV/Wuhan/WIV04/2019), in both Vero and Huh7 cells using BALF sample from ICU-06 patient. Clear cytopathogenic effects were observed in cells after three days incubation (Extended Data Figure 6a and 6b). The identity of the strain WIV04 was verified in Vero E6 cells by immunofluorescence microscopy using cross-reactive viral NP antibody (Extended Data Figure 6c and 6d), and by metagenomic sequencing, from which most of the reads mapped to 2019-nCoV and qPCR showing viral load increase from day 1 to day 3 (Extended Data Figure 6e and 6f). Viral particles in ultrathin sections of infected cells displayed typical coronavirus morphology under electron microscopy (Extended Data Figure 6g). To further confirm the neutralization activity of the viral IgG positive samples, we conducted serum-neutralization assays in Vero E6 cells using the five IgG positive patient sera. We demonstrate that all samples were able to neutralize 120 TCID50 2019-nCoV at a dilution of 1:40-1:80. We also show that this virus could be cross-neutralized by horse anti-SARS-CoV serum (offered by L-F Wang) at dilutions 1:80, but the potential for cross reactivity with SARS-CoV antibodies needs to be confirmed with anti-SARS-CoV serum from humans (Extended Data Table 4).

Angiotensin converting enzyme II (ACE2) was known as cell receptor for SARS-CoV[14]. To determine whether 2019-nCoV also use ACE2 as a cellular entry receptor, we conducted virus infectivity studies using HeLa cells expressing or not expressing ACE2 proteins from humans, Chinese horseshoe bats, civet, pig, and mouse. We show that 2019-nCoV is able to use all but mouse ACE2 as an entry receptor in the ACE2-expressing cells, but not cells without ACE2, indicating which is likely the cell receptor of 2019-nCoV (Fig. 3). We also proved that 2019-nCoV does not use other coronavirus receptors, aminopeptidase N and dipeptidyl peptidase 4 (Extended Data Figure 7).

The study provides the first detailed report on 2019-nCoV, the likely etiology agent responsible for ongoing acute respiratory syndrome epidemic in Wuhan, central China. Viral specific nucleotide positive and viral protein seroconversion observed in all patients tested provides evidence of an association between the disease and the presence of this virus. However, there are still many urgent questions to be answered. The association between the 2019-nCoV and the disease has not been proved by animal experiments to full the Koch postulates. We don't know the transmission routine of this virus among hosts yet. It seems the virus is becoming more transmissible between human-to-human. We should closely monitor if the virus continue evolving to become more virulent. Owing to shortage of specific treatment and considering the relatedness between SARS-CoV and 2019-nCoV, some drugs and pre-clinical vaccine against SARS-CoV probably can be applied to this virus. Finally, considering the wide spread of SARSr-CoV in their natural reservoirs, future research should be focused on active surveillance of these viruses through a broader geographic regions. In the long-term, broad-spectrum antiviral drugs and vaccine should be prepared for the future emerging infectious diseases caused by this cluster of virus. Most importantly, strict regulations against the wildlife domestication and consuming should be implemented.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2012-7.

1. Li, W. D. *et al*. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676-679, (2005).
2. Ge, X. Y. *et al*. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-+, (2013).
3. Yang, L. *et al*. Novel SARS-like Betacoronaviruses in Bats, China, 2011. *Emerg Infect Dis* **19**, 989-991, (2013).
4. Hu, B. *et al*. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS pathogens* **13**, e1006698, (2017).
5. Menachery, V. D. *et al*. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* **21**, 1508-1513, (2015).
6. Menachery, V. D. *et al*. SARS-like WIV1-CoV poised for human emergence. *Proc Natl Acad Sci U S A* **113**, 3048-3053, (2016).
7. Wang, N. *et al*. Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virol Sin* **33**, 104-107, (2018).
8. Drosten, C. *et al*. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New Engl J Med* **348**, 1967-1976, (2003).

9. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. M. E. & Fouchier, R. A. M. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *New Engl J Med* **367**, 1814-1820, (2012).
10. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192, (2019).
11. Fan, Y., Zhao, K., Shi, Z. L. & Zhou, P. Bat Coronaviruses in China. *Viruses* **11**, (2019).
12. Wuhan Municipal Health Commission. *Press statement related to novel coronavirus infection (in Chinese)*, 2020).
13. Poon, L. L. *et al*. Identification of a novel coronavirus in bats. *Journal of virology* **79**, 2001-2009, (2005).
14. Li, W. *et al*. Angiotensin-Converting Enzyme 2 Is a Functional Receptor for the Sars Coronavirus. *Nature* **426**, 450-454, (2003).

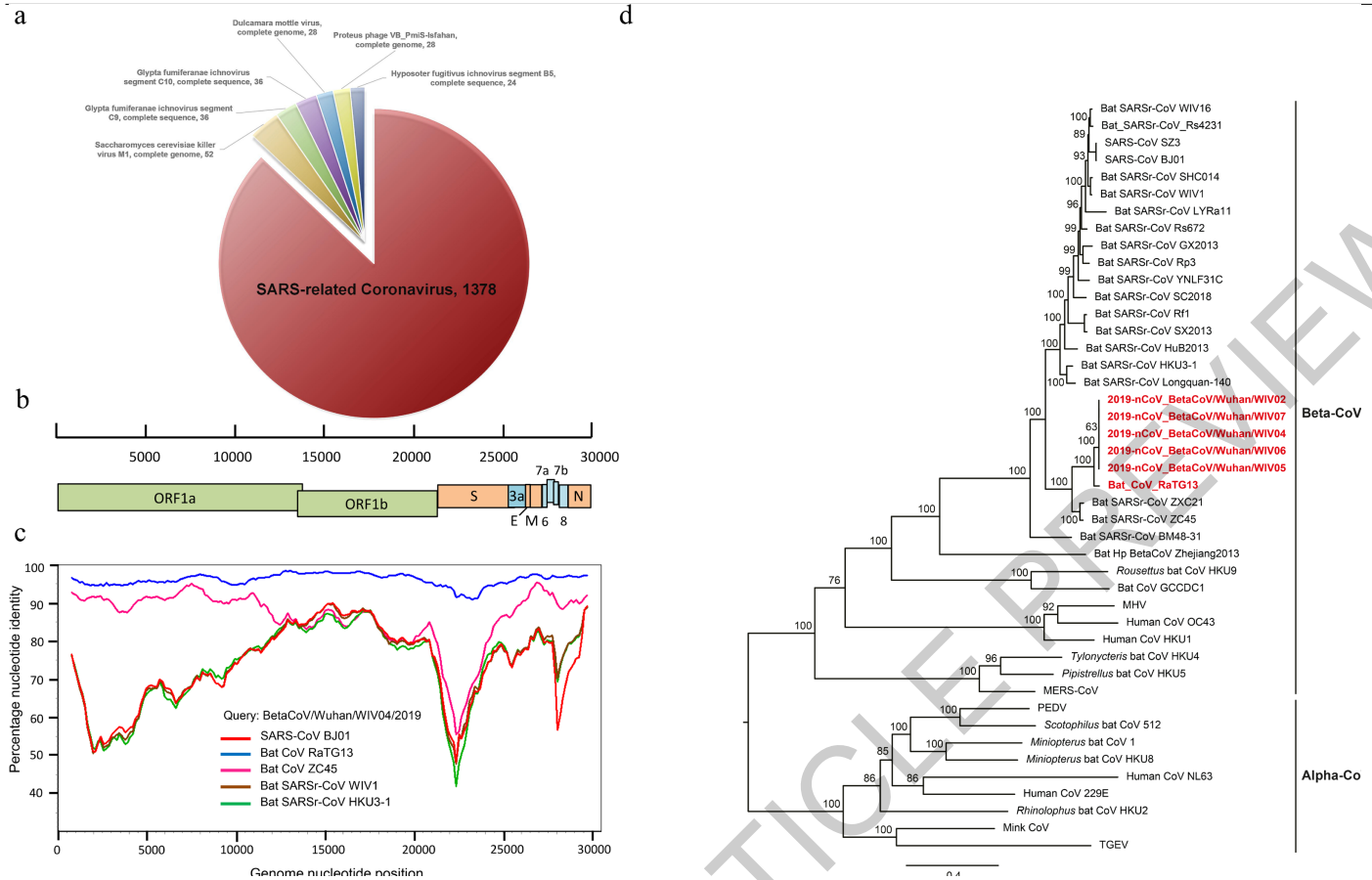


Fig. 1 | Genome characterization of 2019-nCoV. a, pie chart showing metagenomics analysis of next-generation sequencing of bronchoalveolar lavage fluid from patient ICU06. **b**, Genomic organization of 2019-nCoV WIV04. **c**, Similarity plot based on the full-length genome sequence of 2019-nCoV WIV04. Full-length genome sequences of SARS-CoV BJ01, bat SARSr-CoV WIV1, bat coronavirus RaTG13 and ZC45 were used as reference sequences. **d**, Phylogenetic tree based on nucleotide sequences of complete genomes of coronaviruses. Software used and settings can be found in material and method section.

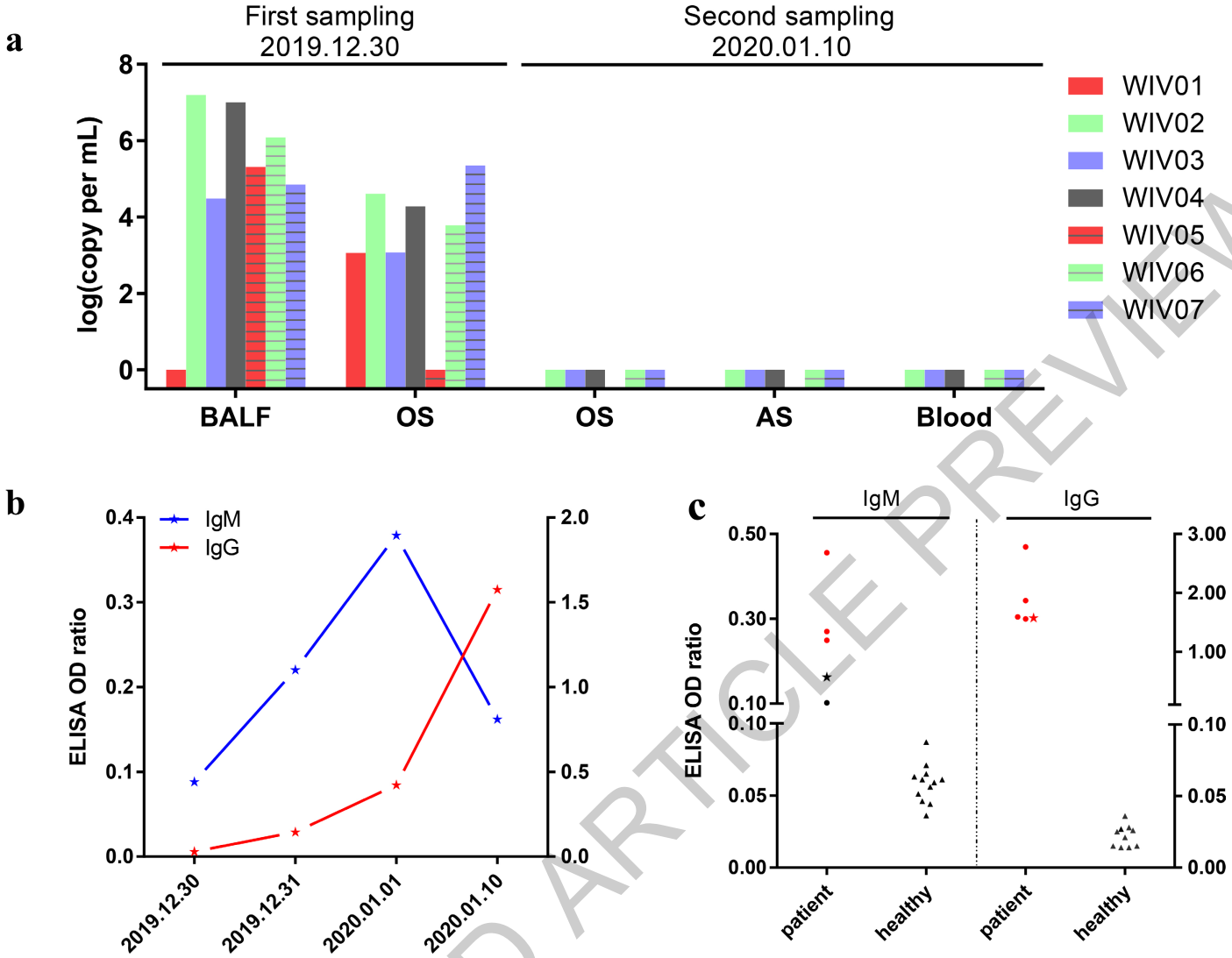


Fig. 2 | Molecular and serological investigation of patient samples. a, molecular detection of 2019-nCoV in seven patients during two times of sampling. Patient information can be found in Extended Data Table 1 and 2. Details on detection method can be found in material and methods. BALF, bronchoalveolar lavage fluid; OS, oral swab; AS, anal swab. **b**, dynamics of 2019-nCoV antibodies in one patient who showed sign of disease on 2019.12.23 (ICU-06). **c**, serological test of 2019-nCoV antibodies in five patients (more information can be found in Extended Data Table 2). Star indicates data collected from patient ICU-06 on 2020.01.10. For b and c, cut-off was set up as 0.2 for IgM test and 0.3 for IgG test, according to healthy controls.

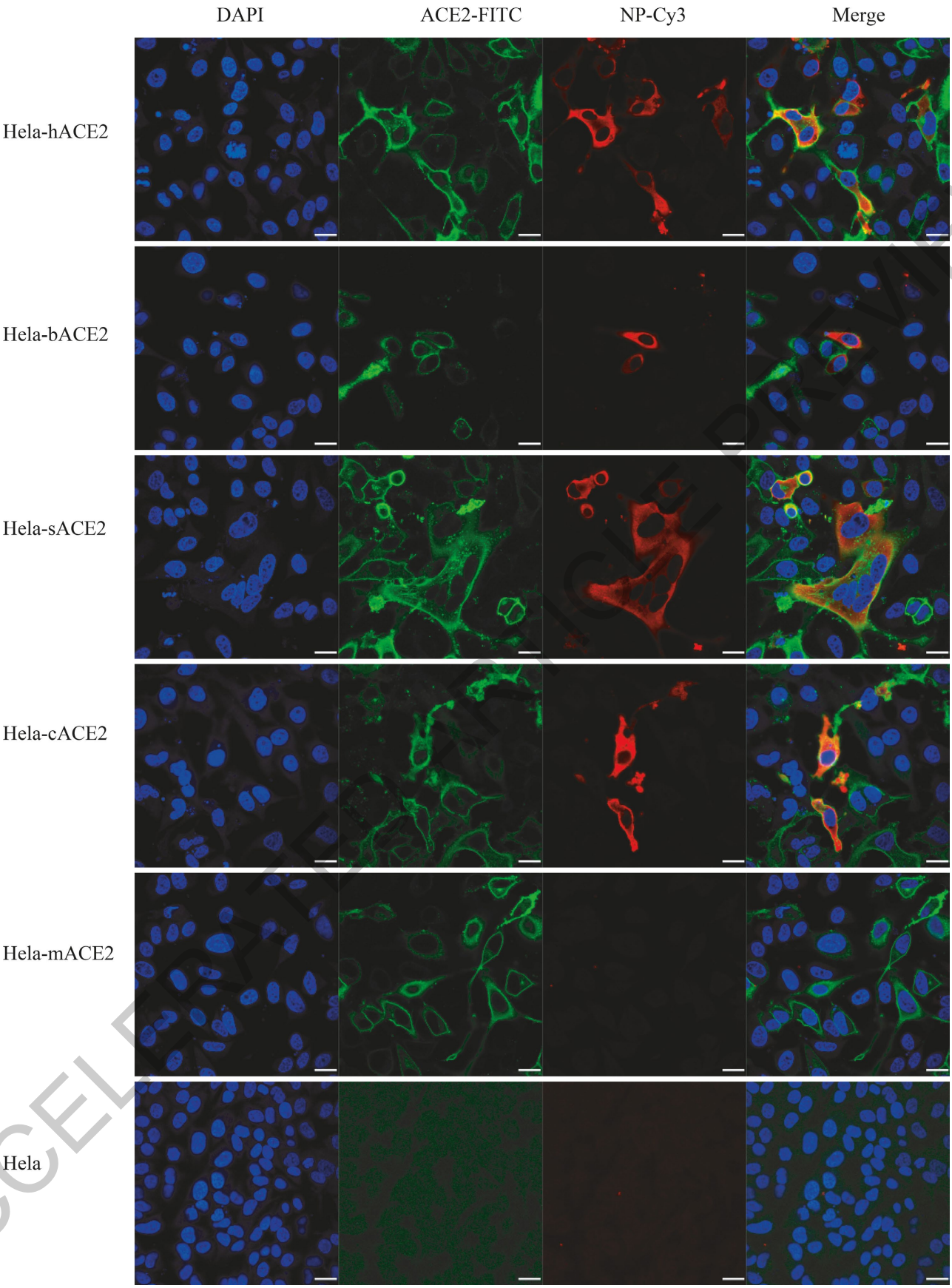


Fig. 3 | Analysis of 2019-nCoV receptor usage. Determination of virus infectivity in HeLa cells with or without the expression of ACE2. h, human; b, *Rhinolophus sinicus* bat; c, civet; s, swine (pig); m, mouse. ACE2 protein (green), viral protein (red) and nuclei (blue) was shown. Scale bar=10 um.

Methods

Sample collection

Human samples, including oral swabs, anal swabs, blood, and BALF samples were collected by Jinyintan hospital (Wuhan) with the consent from all patients and approved by the ethics commission of the designated hospital for emerging infectious diseases. Patients were sampled without gender or age preference unless where indicated. For swabs, 1.5 ml DMEM+2% FBS medium was added each tube. Supernatant was collected after 2500 rpm, 60 s vortex and 15-30 min standing. Supernatant from swabs or BALF (no pretreatment) was added to either lysis buffer for RNA extraction or to viral transport medium (VTM) for virus isolation. VTM composed of Hank's balanced salt solution at pH7.4 containing BSA (1%), amphotericin (15 μg/ml), penicillin G (100 units/ml), and streptomycin (50 μg/ml). Serum was separated by centrifugation at 3,000 *g* for 15 min within 24 h of collection, followed by 56 ^{0}C 30 min inactivation, and then stored at 4 ^{0}C until use.

Virus isolation, cell infection, electron microscope and neutralization assay

The following cells were used for virus isolation in this study: Vero, Vero E6, and Huh7 that were cultured in DMEM +10% FBS. A list of cells were used for susceptibility test (Extended Data Fig. 6). All cell lines were tested free of mycoplasma contamination, applied to species identification and authenticated by microscopic morphologic evaluation. None of cell lines was on the list of commonly misidentified cell lines (by ICLAC).

Cultured cell monolayers were maintained in their respective medium. PCR-positive BALF sample from ICU-06 patient was spin at 8,000 g for 15 min, filtered and diluted 1:2 with DMEM supplied with 16 μg/ml trypsin before adding to cells. After incubation at 37 ^{0}C for 1 h, the inoculum was removed and replaced with fresh culture medium containing antibiotics (below) and 16 μg/ml trypsin. The cells were incubated at 37 ^{0}C and observed daily for cytopathic effect (CPE). The culture supernatant was examined for presence of virus by qRT-PCR developed in this study, and cells were examined by immunofluorescent using SARSr-CoV Rp3 NP antibody made in house (1:100). Penicillin (100 units/ml) and streptomycin (15 μg/ml) were included in all tissue culture media.

The Vero E6 cells were infected with new virus at MOI of 0.5 and harvested 48 hpi. Cells were fixed with 2.5% (wt/vol) glutaraldehyde and 1% osmium tetroxide, and then dehydrated through a graded series of ethanol concentrations (from 30 to 100%), and embedded with epoxy resin. Ultrathin sections (80 nm) of embedded cells were prepared, deposited onto Formvar-coated copper grids (200 mesh), stained with uranyl acetate and lead citrate, then observed under 200 kV Tecnai G2 electron microscope.

The virus neutralization test was carried out in a 48-well plate. The patient serum samples were heat-inactivated by incubation at 56 ^{0}C for 30 min before use. The serum samples (5 μL) were diluted to 1:10, 1:20, 1:40 or 1:80, and then an equal volume of virus stock was added and incubated at 37 ^{0}C for 60 min in a 5% CO_2 incubator. Diluted horse anti SARS-CoV serum or serum samples from healthy people were used as control. After incubation, 100 μL mixtures were inoculated onto monolayer Vero E6 cells in a 48-well plate for 1 hour. Each serum were repeated triplicate. After removing the supernatant, the plate was washed twice with DMEM medium. Cells were incubated with DMEM supplemented with 2% FBS for 24 hours. Then the cells were fixed with 4% formaldehyde. And the virus were detected using SL-CoV Rp3 NP antibody followed by Cy3-conjugated mouse anti-rabbit IgG. Nuclei were stained with DAPI. Infected cell number was counted by high-content cytometers.

RNA extraction and PCR

Whenever commercial kits were used, manufacturer's instructions were followed without modification. RNA was extracted from 200 μl of samples with the High Pure Viral RNA Kit (Roche). RNA was eluted in 50 μl of elution buffer and used as the template for RT-PCR.

For qPCR analysis, primers based on 2019-nCoV S gene was designed: RBD-qF1: 5'-CAATGGTTTAACAGGCACAGG-3'; RBD-qR1: 5'-CTCAAGTGTCTGTGGATCACG-3'. RNA extracted from above used in qPCR by HiScript® II One Step qRT-PCR SYBR® Green Kit (Vazyme Biotech Co.,Ltd). Conventional PCR test was also performed using the following primer pairs: ND-CoVs-951F TGTKAGRTTYCCTAAYATTAC; ND-CoVs-1805R ACATCYTGATANARAACAGC[13]. The 20 μl qPCR reaction mix contained 10 μl 2× One Step SYBR Green Mix, 1 μl One Step SYBR Green Enzyme Mix, 0.4 μl 50 × ROX Reference Dye 1, 0.4 μl of each primer (10 uM) and 2 μl template RNA. Amplification was performed as follows: 50 ^{0}C for 3 min, 95 ^{0}C for 30 s followed by 40 cycles consisting of 95 ^{0}C for 10 s, 60 ^{0}C for 30 s, and a default melting curve step in an ABI 7700 machine.

Serological test

In-house anti-SARSr-CoV IgG and IgM ELISA kits were developed using SARSr-CoV Rp3 NP as antigen, which shared above 90% amino acid identity to all SARSr-CoVs[2]. For IgG test, MaxiSorp Nunc-immuno 96 well ELISA plates were coated (100 ng/well) overnight with recombinant NP. Human sera were used at 1:20 dilution for 1 h at 37 ^{0}C. An anti-Human IgG-HRP conjugated monoclonal antibody (Kyab Biotech Co., Ltd, Wuhan, China) was used at a dilution of 1:40000. The OD value (450−630) was calculated. For IgM test, MaxiSorp Nunc-immuno 96 wellELISA plates were coated (500 ng/well) overnight with anti-human IgM (μ chain). Human sera were used at 1:100 dilution for 40 min at 37 ^{0}C, followed by anti-Rp3 NP-HRP conjugated (Kyab Biotech Co., Ltd, Wuhan, China) at a dilution of 1:4000. The OD value (450−630) was calculated.

Examination of ACE2 receptor for 2019-nCoV infection

HeLa cells transiently expressing ACE2 were prepared by a lipofectamine 3000 system (Thermo Fisher Scientific) in 96-well plate, with mock-transfected cells as controls. 2019-nCoV grown from Vero E6 cells was used for infection at multiplicity of infection 0.05. Same for testing of APN and DPP4. The inoculum was removed after 1 h absorption and washed twice with PBS and supplemented with medium. At 24 hpi, cells were washed with PBS and fixed with 4% formaldehyde in PBS (pH 7.4) for 20 min at room temperature. ACE2 expression was detected using mouse anti-S tag monoclonal antibody followed by FITC-labelled goat anti-mouse IgG H&L (Abcam, ab96879). Viral replication was detected using rabbit antibody against the Rp3 NP protein (made in house, 1:100) followed by cyanin 3-conjugated goat anti-rabbit IgG (1:50, Abcam, ab6939). Nucleus was stained with DAPI (Beyotime). Staining patterns were examined using the FV1200 confocal microscopy (Olympus).

High throughput sequencing, pathogen screening and genome assembly

Samples from patient BALF or from virus culture supernatant were used for RNA extraction and next-generation sequencing using BGI MGISEQ2000 and Illumina MiSeq 3000 sequencers. Metagenomic analysis was carried out mainly base on the bioinformatics platform MGmapper (PE_2.24 and SE_2.24). The raw NGS reads were firstly processed by Cutadapt (v1.18) with minimum read length of 30bp. BWA (v0.7.12-r1039) was utilized to align reads to local database with a filter hits parameter at 0.8 FMM value and minimum alignment score at 30. Parameters for post-processing of assigned reads was set with minimum size normalized abundance at 0.01, minimum read count at 20 and other default parameters. A local nucleic acid database for human and mammals was employed to filter reads of host genomes before mapping reads to virus database. The results of metagenomic analysis were displayed through pie charts using WPS Office 2010. NGS reads were assembled into genomes using Geneious (v11.0.3) and

MEGAHIT (v1.2.9). PCR and Sanger sequencing was performed to fill gaps in the genome. 5′-RACE was performed to determine the 5′-end of the genomes using SMARTer RACE 5′/3′ Kit (Takara). Genomes were annotated using Clone Manager Professional Suite 8 (Sci-Ed Software).

Phylogenetic analysis

Routine sequence management and analysis was carried out using DNAStar. The sequence alignment of complete genome sequences was performed by MAFFT (version 7.307) with default parameters. The codon alignments of full-length S and RdRp gene sequences were converted from the corresponding protein alignments by PAL2NAL (version 14), respectively, of which the protein alignments were created by Clustal Omega (version 1.2.4) under default parameters. Maximum Likelihood phylogenetic trees were carried out using RAxML (version 0.9.0) with GTR+G substitution model and 1000 bootstrap replicates.

Data availability

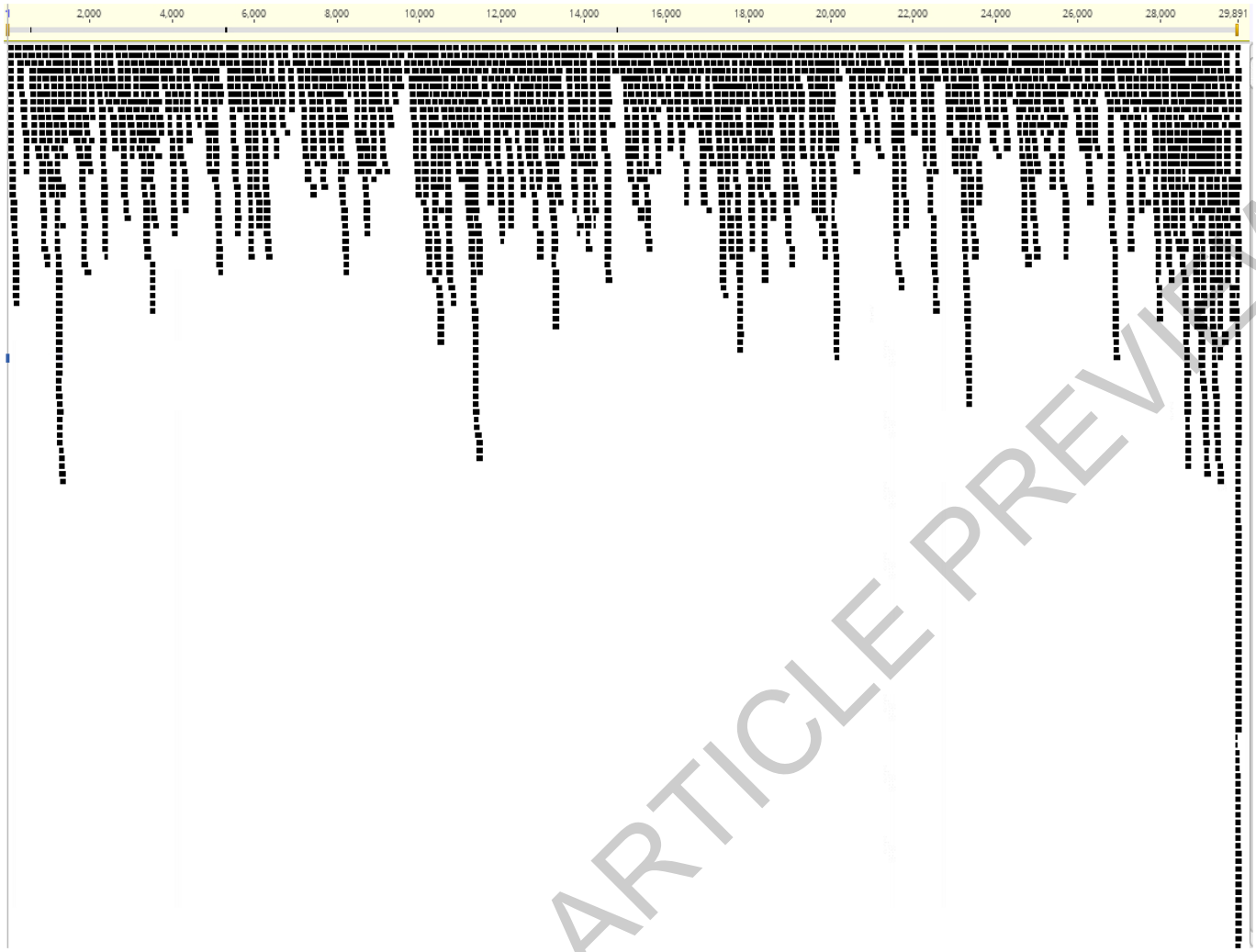
Sequence data that support the findings of this study have been deposited in GISAID with the accession numbers EPI_ISL_402124, EPI_ISL_402127–EPI_ISL_402130 and EPI_ISL_402131.

Acknowledgements We thank the Pei Zhang and An-na Du from WIV core facility and technical support for their help with producing TEM micrographs. We thank Hai-Zhou Liu and Ping Yu from WIV for bioinformatics analysis. This work was jointly supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB29010101 to ZLS and XDB29010104 to PZ), China Natural Science Foundation for excellent scholars (81822028 to PZ, 31770175 to ZLS and 31800142 to BH), Mega-Project for Infectious Disease from Minister of Science and Technology of the People's Republic of China (2020ZX09201001 to DYZ and 2018ZX10305409-004-001 to PZ), Youth innovation promotion association of CAS (2019328 to XLY).

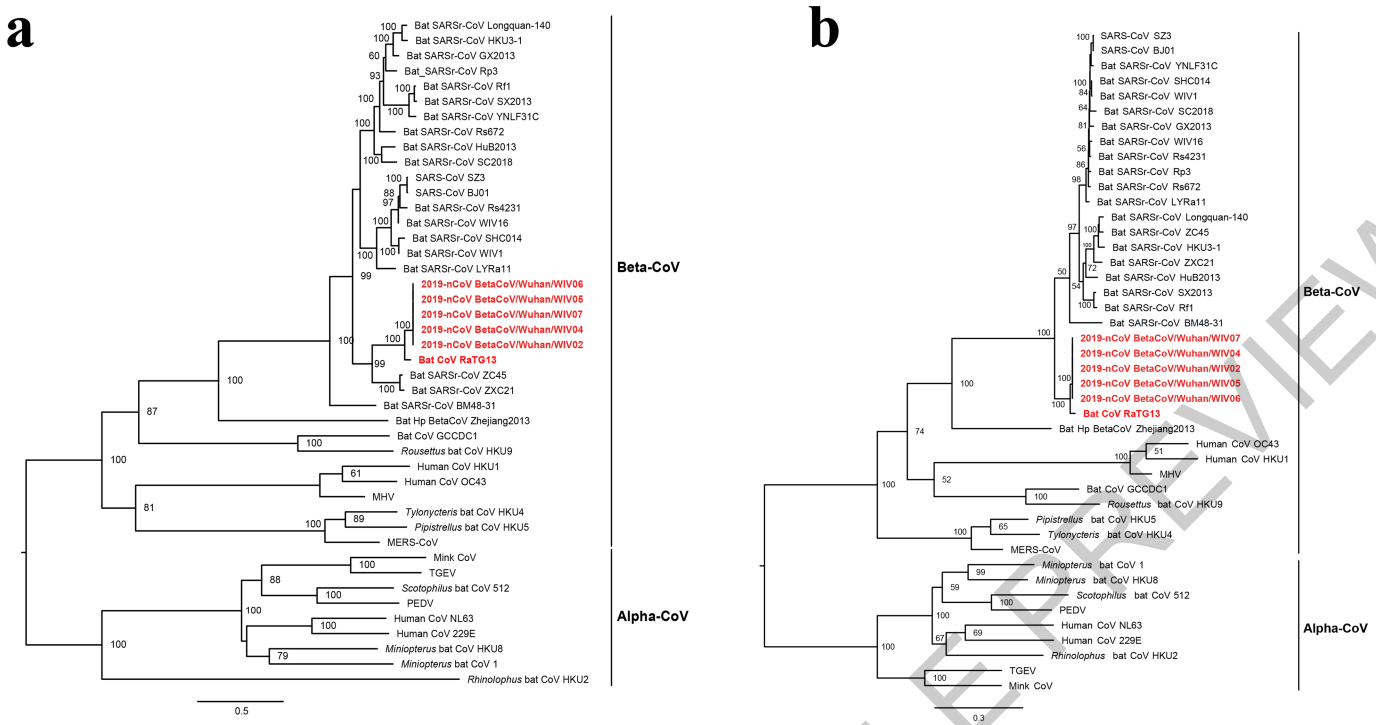
Author contributions Z.L.S., P.Z., Y.Y.W., and G.F.X. conceived the study. G.S.W., C.L.H., H.D.C., F.D., Q.J.C., F.X.Z., and LLL., collected patient samples. X.L.Y., B.Y., W.Z., B.L., J.C., X.S.Z., Y.L., H.G., R.D.J., M.Q.L., Y. Chen, X.W., X.R.S., and K.Z. performed qPCR, serology, and virus culturing. L.Z., Y.Z., H.R.S., and B.H. performed genome sequencing and annotations.

Competing interests The authors declare no competing interests.

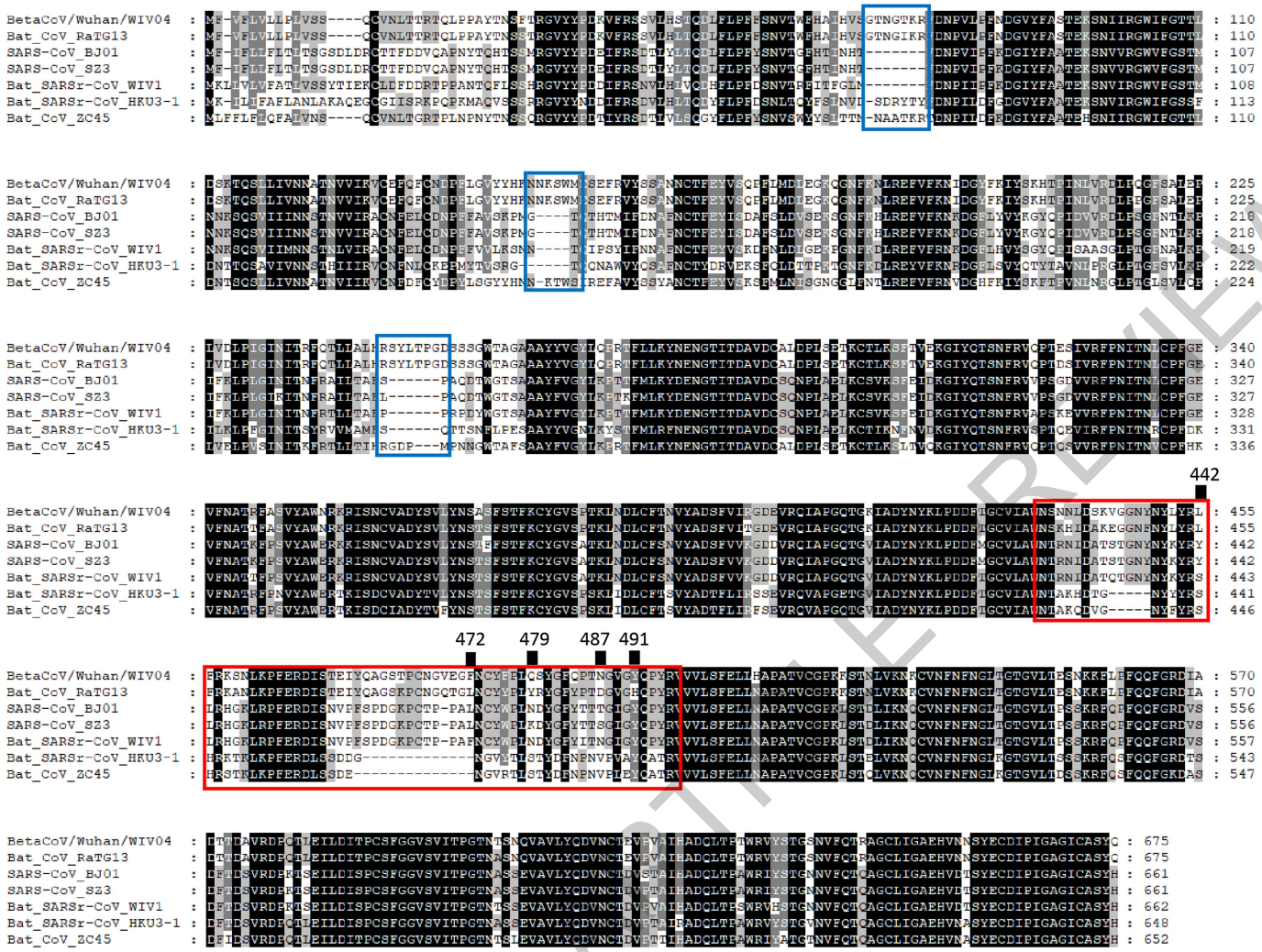
Additional information
Correspondence and requests for materials should be addressed to Z.-L.S.
Reprints and permissions information is available at http://www.nature.com/reprints.



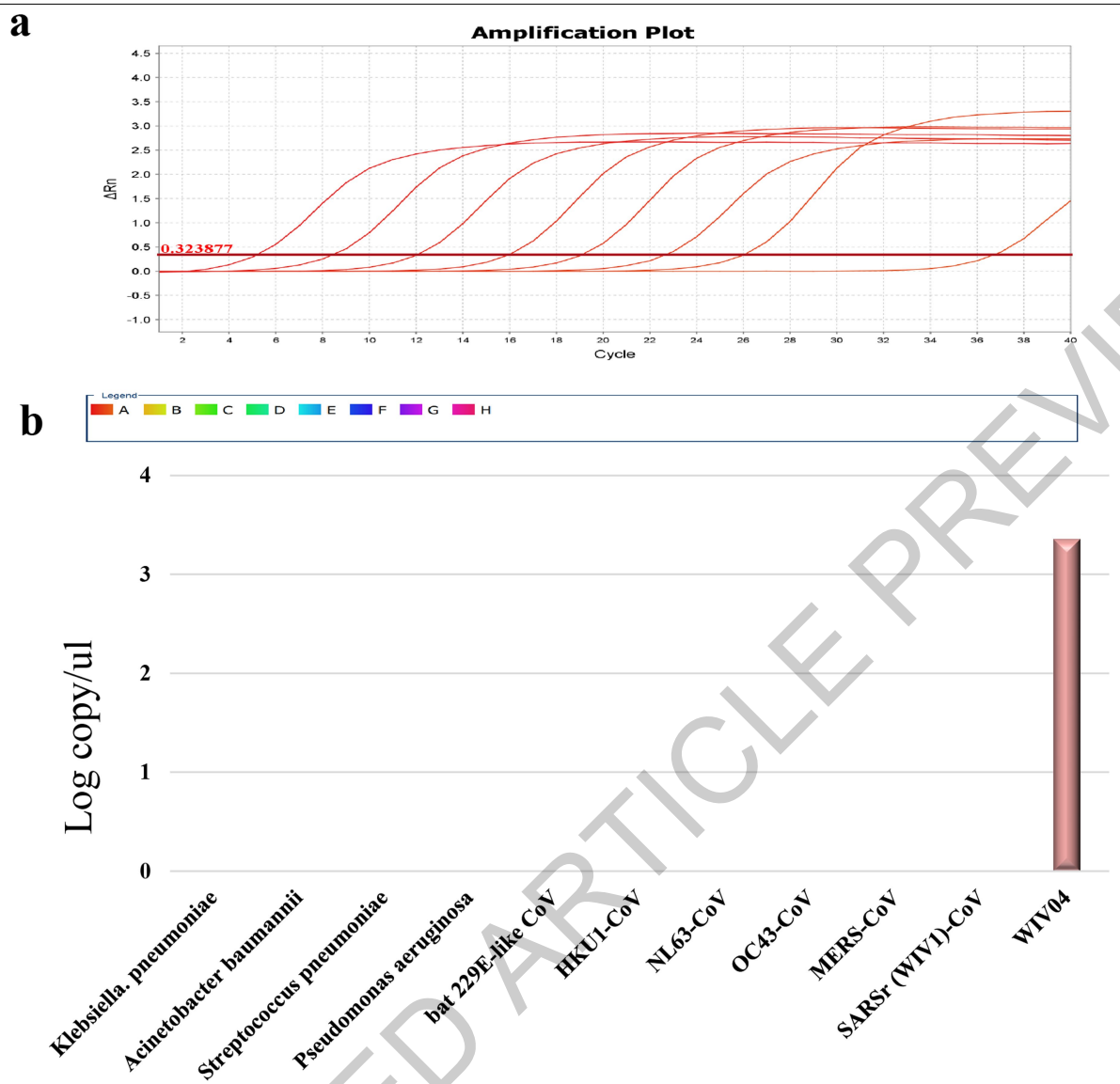
Extended Data Fig. 1 | NGS raw reads from WIV04 patient mapping to 2019-nCoV.



Extended Data Fig. 2 | Phylogenetic tree base on the complete S (a) and RdRp (b) gene sequences of coronaviruses. 2019-nCoV and bat CoV RaTG13 are shown in bold and in red. The trees were constructed by the maximum likelihood method using the GTR+G substitution model with bootstrap values determined by 1000 replicates. Bootstraps > 50% are shown.

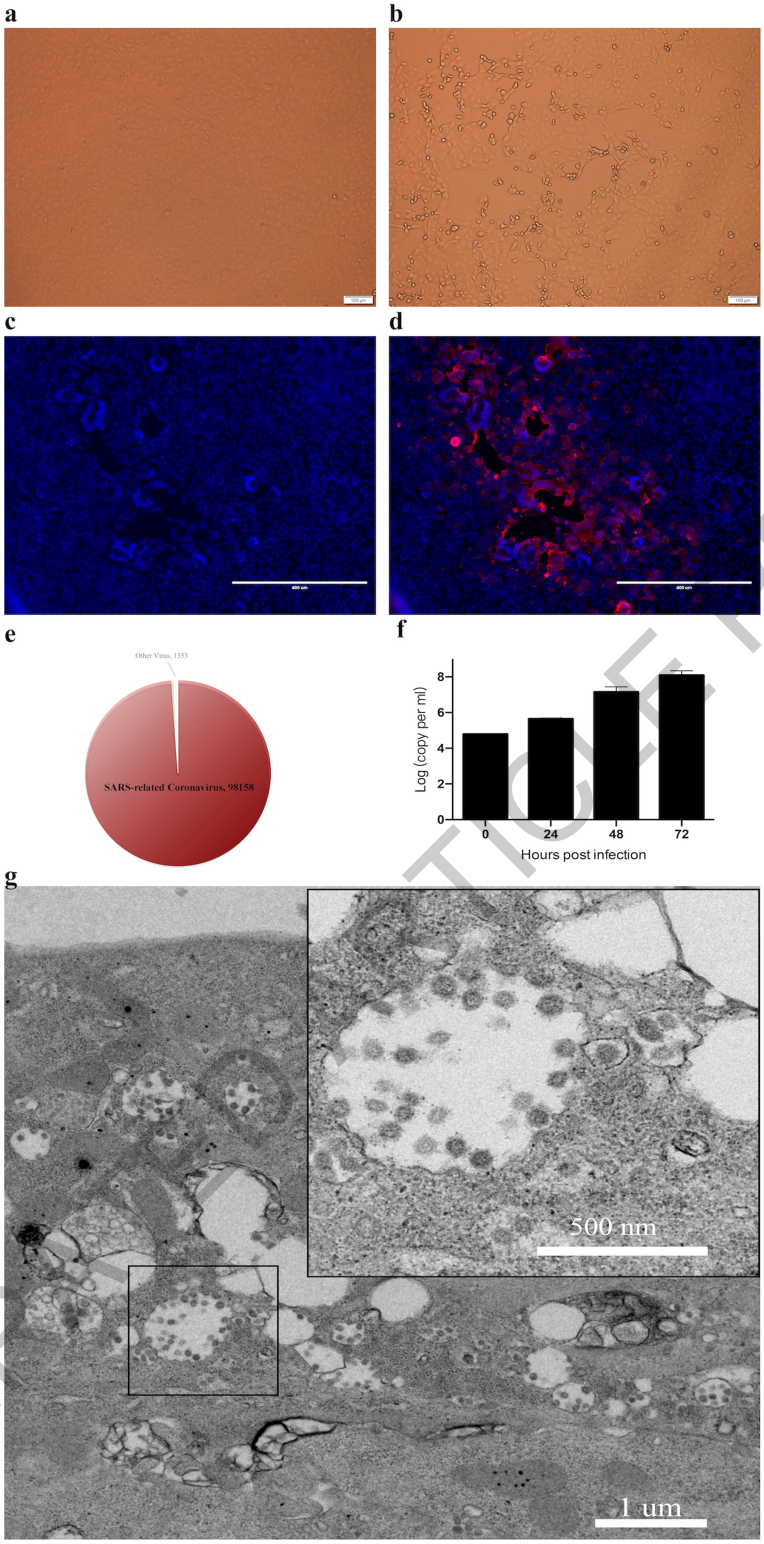


Extended Data Fig. 3 | Amino acid sequence alignment of the S1 protein of the 2019-nCoV with SARS-CoV and selected bat SARSr-CoVs. The receptor-binding motif of SARS-CoV and the homologous region of other coronaviruses are indicated by the red box. The key amino acid residues involved in the interaction with human ACE2 are numbered on top of the aligned sequences. The short insertions in the N-terminal domain of the novel coronavirus are indicated by the blue boxes. Bat CoV RaTG13 was identified from *R. affinis* in Yunnan Province. Bat CoV ZC45 was identified from *R. sinicus* in Zhejiang Province.

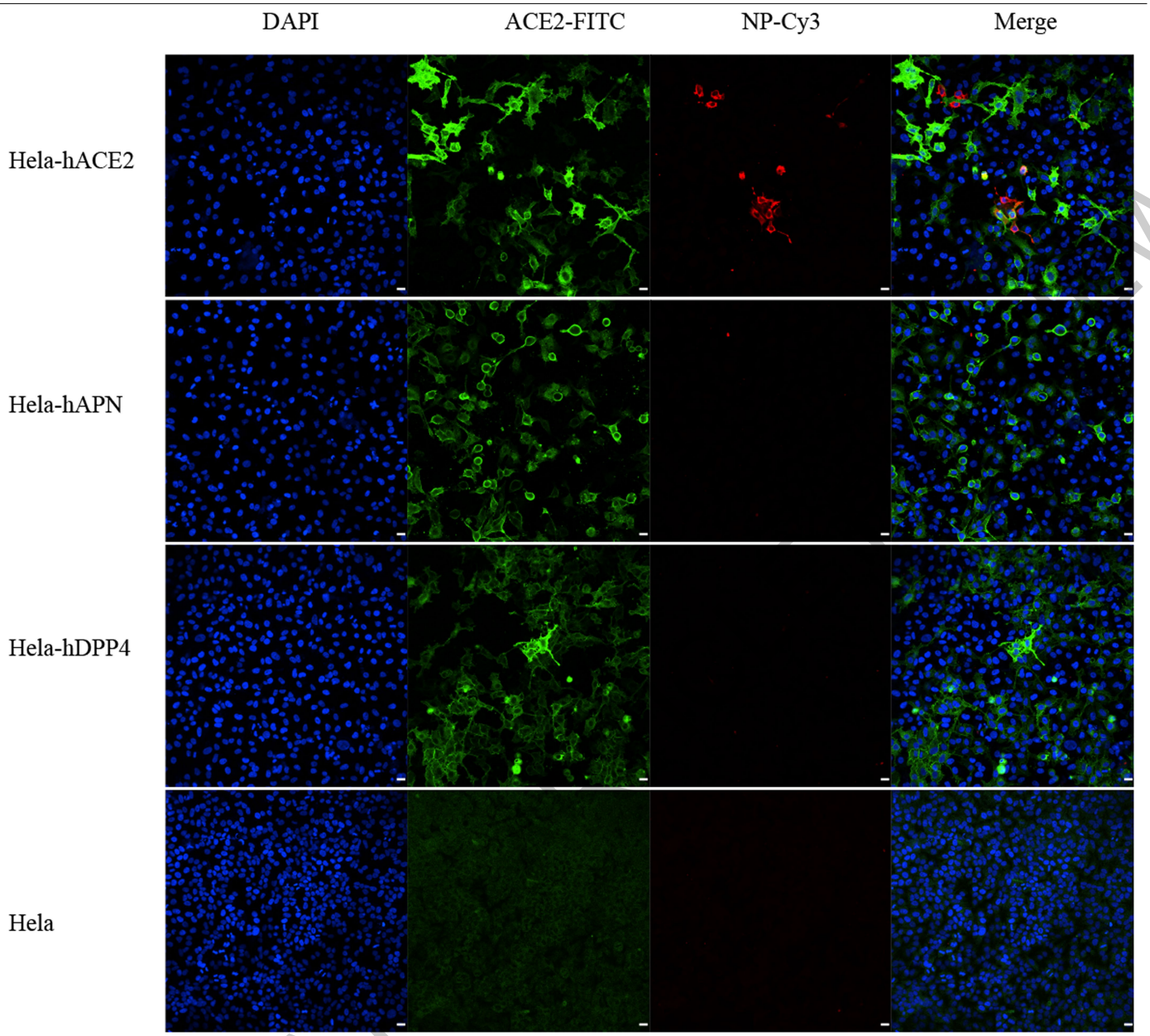


Extended Data Fig. 4 | Molecular detection method set up for 2019-nCoV. a, standard curve for qPCR primers. PCR product of spike gene that was serial diluted to 10^8 to 10^1 (from left to right) was used as template. Primer sequence and experiment condition can be found in material and methods. **b**, specificity of qPCR primers. Nucleotide samples from the indicated pathogens were used.

Extended Data Fig. 5 | Amino acid sequence alignment of the nucleocapsid protein of 2019-nCoV with bat SARSr-CoV Rp3 and SARS-CoV BJ01.



Extended Data Fig. 6 | Isolation and antigenic characterization of 2019-nCoV. Vero E6 cells are shown at 24 hours post infection with mock (**a**) or 2019-nCoV (**b**). (**c**) and (**d**) are mock or 2019-nCoV infected samples stained with rabbit serum raised against recombinant SARSr-CoV Rp3 N protein (red) and DAPI (blue). The experiment was conducted two times independently with similar results. **e** and **f**, pie charts illustrating ratio of reads number related to 2019-nCoV among total viral related reads in metagenomics analysis of Vero (**e**) and Huh7 (**f**) cell culture supernatant. (**g**) viral particles in the ultrathin sections under electron microscope at 200 kV, sample from viral infected Vero E6 cells.



Extended Data Fig. 7 | Analysis of 2019-nCoV receptor usage. Determination of virus infectivity in HeLa cells with or without the expression of human APN and DPP4. ACE2 protein (green), viral protein (red) and nuclei (blue) were shown. Scale bar=10 um.

Extended Data Table 1 | Patient information and their diagnosis history (some records are missing)

Patient No.	Gender	Age	Date of Onset	Date of Admission	Symptoms When Admitted	Current Status (2020.01.13)	Diagnosis history
ICU-01*	Male	62	2019.12.12	2019.12.27	fever	recover, discharged	negative
ICU-04	Male	32	2019.12.19	2019.12.29	fever, cough, dyspnea	fever, intermittent cough	negative
ICU-05	Male	40	2019.12.17	2019.12.27	fever (38 °C), expectoration, malaise, dyspnea	fever, malaise, intermittent cough	AdV (IgM)
ICU-06	Female	49	2019.12.23	2019.12.27	fever (37.9 °C), palpitation	fever, malaise, cough	Coronavirus (nt)
ICU-08	Female	52	2019.12.22	2019.12.29	fever (38.5 °C), expectoration, malaise, dyspnea	recover, discharged	Streptococcus pneumoniae (nt)
ICU-09	Male	40	2019.12.22	2019.12.28	fever (38.5 °C), expectoration	fever (38.5 °C), malaise, expectoration, dizziness	negative
ICU-10	Male	56	2019.12.20	2019.12.20	fever, dyspnea, chest tightness	fever, malaise, cough, dyspnea	negative

All patients are seafood market sellers or deliverymen except ICU-01, whose contact history is unclear. All patients were in intensive care unit (ICU) during the first investigation, and now in stable condition. Blood IgM tests have been performed for the following respiratory pathogens for all patients: legionella pneumophilia, mycoplasma pneumoniae, chlamydia pneumoniae, respiratory syncytial virus, adenovirus, rickettsia, influenza A virus, influenza B virus, parainfluenza virus. *This patient reported fever on 2019.12.12, and then recovered without medical treatment. He came back to hospital on 2019.12.27 due to fever. His wife was also sick and admitted to hospital. Both of them were recovered.

Extended Data Table 2 | Laboratory detection results

		First sampling-2019.12.30			Second sampling-2020.01.10			
Patient No.	Test No.	BALF	Oral Swab	Blood (Ab)	Oral Swab	Anal Swab	Blood (PCR)	Blood (Ab)
ICU-01	WIV01	-	Ct=32.0	NA	NA	NA	NA	NA
ICU-04	WIV02[#]	Ct=17.6	Ct=26.6	NA	-	-	-	+
ICU-05	WIV03	Ct=27.0	Ct=31.9	NA	-	-	-	+
ICU-06	WIV04[#*]	Ct=18.3	Ct=27.7	+	-	-	-	+
ICU-08	WIV05[#]	Ct=24.1	-	NA	NA	NA	NA	NA
ICU-09	WIV06[#]	Ct=21.6	Ct=29.4	NA	-	-	-	+
ICU-10	WIV07[#]	Ct=25.7	Ct=24.0	NA	-	-	-	+

Samples from two patients (ICU-01 and ICU-08) were not available during the second investigation. They have been discharged from hospital. We did serial test for ICU-06 patient at the following date: 19.12.30, 19.12.31, 20.01.01 and 20.01.10, corresponding to seven, eight, nine and eighteen days upon disease onset (19.12.23). Table shows molecular and serological (IgM and IgG) detection results for 2019-nCoV. [$]Full-length genome obtained. *Virus isolated.

Extended Data Table 3 | Genomic comparison of 2019-nCoV WIV04 with SARS-CoVs and bat SARSr-CoVs

	Sequence identities with SARS-CoVs & bat SARSr-CoVs (nt/aa %)											
	Full-length genome	ORF1a	ORF1b	S	ORF3a	E	M	ORF6	ORF7a	ORF7b	ORF8	N
SARS-CoV GZ02	79.6	76.0/80.9	86.2/95.7	73.4/77.0	75.6/73.4	94.7/96.0	85.4/90.5	76.3/68.9	82.8/86.0	84.8/81.4	52.0/31.6	87.7/91.2
SARS-CoV BJ01	79.6	76.0/80.8	86.2/95.7	73.4/76.9	75.3/72.6	94.7/96.0	85.6/90.5	75.8/67.2	82.8/86.0	84.8/81.4	51.1/-	88.8/91.2
SARS-CoV Tor2	79.6	76.0/80.9	86.2/95.8	73.4/76.7	75.4/72.6	94.7/96.0	85.6/90.5	76.3/68.9	82.8/86.0	84.8/81.4	51.1/-	88.8/91.2
SARS-CoV SZ3	79.6	76.0/81.0	86.2/95.8	73.4/76.9	75.4/72.6	94.7/96.0	85.3/90.0	76.3/68.9	82.8/86.0	84.8/81.4	52.3/31.6	88.8/91.2
SARS-CoV PC4-227	79.5	76.0/80.8	86.1/95.6	73.4/76.7	75.5/72.6	94.7/96.0	85.1/90.0	75.8/68.9	82.8/86.0	84.8/81.4	52.3/-	88.5/90.7
Bat SARr-CoV RaTG13	96.2	96.0/98.0	97.3/99.3	93.1/97.7	96.3/97.8	99.6/100	95.5/99.6	98.4/100	95.6/97.5	99.2/97.7	97.0/95.0	96.9/99.0
Bat SARr-CoV WIV1	79.7	76.0/80.7	85.9/95.8	73.4/77.6	76.1/74.5	95.6/96.0	84.8/90.0	78.0/73.8	85.0/88.4	85.6/83.7	65.8/57.9	88.5/90.9
Bat SARSr-CoV WIV16	79.7	75.9/81.0	86.1/95.6	73.1/77.8	76.1/74.5	95.6/96.0	84.8/90.0	77.4/72.1	85.0/88.4	85.6/83.7	65.3/57.9	88.6/90.9
Bat SARSr-CoV SHC014	79.6	75.9/80.9	85.9/95.8	73.3/77.7	76.1/74.5	95.6/96.0	84.8/90.0	78.0/70.5	84.4/88.4	85.6/83.7	65.8/58.7	88.6/90.9
Bat SARSr-CoV Rs4231	79.7	76.0/81.0	86.2/95.8	72.9/77.5	75.8/74.1	94.3/94.7	84.4/90.0	76.9/67.2	85.0/88.4	85.6/83.7	65.3/57.9	88.8/91.4
Bat SARSr-CoV YNLF31C	79.0	75.7/80.6	85.8/95.7	71.4/75.5	75.0/71.2	94.3/96.0	84.7/89.6	76.9/70.5	83.1/87.6	86.4/83.7	50.3/31.3	88.3/90.5
Bat SARSr-CoV LYRa11	79.6	75.8/80.6	85.7/95.6	73.9/77.3	77.2/76.3	94.7/94.7	85.1/90.0	78.5/70.5	82.0/85.1	81.1/81.4	66.7/57.9	89.0/91.6
Bat SARSr-CoV ZC45	88.1	91.0/95.7	86.1/96.0	77.8/82.3	87.8/90.9	98.7/100	93.4/98.6	95.2/93.4	88.8/87.6	94.7/93.0	88.5/94.2	91.1/94.3
Bat SARSr-CoV ZXC21	88.0	90.9/95.7	86.2/95.8	77.1/81.7	88.9/92.0	98.7/100	93.4/98.6	95.2/93.4	89.1/88.4	95.5/93.0	88.5/94.2	91.2/94.3
Bat SARSr-CoV HuB2013	79.6	76.3/81.2	85.3/95.7	73.1/76.8	75.4/75.5	95.2/94.7	85.3/91.0	76.3/68.9	84.2/87.6	85.6/83.7	62.0/49.6	88.9/91.6
Bat SARSr-CoV GX2013	79.1	75.9/80.8	86.0/95.9	73.1/77.1	75.6/73.0	94.7/96.0	84.8/91.4	77.4/68.9	85.0/86.8	84.1/79.1	51.4/31.6	87.9/90.2
Bat SARSr-CoV SX2013	78.9	76.2/80.6	85.1/95.5	71.2/75.5	74.7/71.2	94.3/93.3	83.0/89.6	77.4/68.9	84.2/86.8	85.6/83.7	49.7/30.4	86.9/90.2
Bat SARSr-CoV SC2018	79.4	75.8/80.7	85.5/95.2	72.7/76.4	75.0/71.2	94.3/96.0	84.7/90.0	80.0/71.8	85.2/87.6	84.8/83.7	66.1/55.4	88.2/91.2
Bat SARSr-CoV Rs672	79.6	76.0/80.9	85.9/95.8	72.8/76.2	75.2/71.9	95.2/96.0	84.8/89.6	78.5/70.5	84.7/88.4	85.6/83.7	65.8/58.7	87.9/91.2
Bat SARSr-CoV Rp3	79.5	75.9/80.5	86.0/95.7	73.1/77.2	74.9/74.8	95.2/96.0	85.1/90.0	76.9/68.9	83.9/89.3	84.8/83.7	66.4/56.2	88.4/90.7
Bat SARSr-CoV Rf1	78.8	76.2/80.6	84.8/95.3	71.1/75.7	74.3/69.0	94.3/94.7	83.3/89.6	79.0/68.9	84.2/86.8	84.1/83.7	50.6/31.3	86.8/89.5
Bat SARSr-CoV HKU3-1	79.4	76.1/80.9	84.9/95.1	73.4/77.9	75.8/73.4	95.2/96.0	84.7/91.0	75.3/67.2	85.0/89.3	84.1/79.1	66.4/57.0	88.3/90.0

Extended Data Table 4 | Virus neutralization test (VNT) of serum samples

Samples	VNT titre for nCoV-2019
Healthy people #1 from Wuhan	neg
Healthy people #2 from Wuhan	neg
Horse anti-SARS-CoV serum	>1:80
WIV02	>1:80
WIV03	1:40
WIV04	>1:80
WIV06	>1:80
WIV07	>1:80

Each serum sample was tested in triplicate. Two healthy people from Wuhan, five patient serum samples and a horse anti-SARS-CoV anti-serum were used. 120 $TCID_{50}$ viruses were used each well. Serum samples were used in a dilution from 1:10, 1:20, 1:40 to 1:80.