nature

https://doi.org/10.1038/s41586-020-2313-x

Accelerated Article Preview

Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins

Received: 16 February 2020

Accepted: 28 April 2020

Accelerated Article Preview
Published online 7 May 2020

Cite this article as: Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* https://doi.org/10.1038/s41586-020-2313-x (2020).

Kangpeng Xiao, Junqiong Zhai, Yaoyu Feng, Niu Zhou, Xu Zhang, Jie-Jian Zou, Na Li, Yaqiong Guo, Xiaobing Li, Xuejuan Shen, Zhipeng Zhang, Fanfan Shu, Wanyi Huang, Yu Li, Ziding Zhang, Rui-Ai Chen, Ya-Jiang Wu, Shi-Ming Peng, Mian Huang, Wei-Jun Xie, Qin-Hui Cai, Fang-Hui Hou, Wu Chen, Lihua Xiao & Yongyi Shen

Article

Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins

https://doi.org/10.1038/s41586-020-2313-x

Received: 16 February 2020

Accepted: 28 April 2020

Published online: 7 May 2020

Kangpeng Xiao[1,2,7], Junqiong Zhai[3,7], Yaoyu Feng[1,2], Niu Zhou[3], Xu Zhang[1,2], Jie-Jian Zou[5], Na Li[1,2], Yaqiong Guo[1,2], Xiaobing Li[1], Xuejuan Shen[1], Zhipeng Zhang[1], Fanfan Shu[1,2], Wanyi Huang[1,2], Yu Li[4], Ziding Zhang[4], Rui-Ai Chen[1,6], Ya-Jiang Wu[3], Shi-Ming Peng[3], Mian Huang[3], Wei-Jun Xie[3], Qin-Hui Cai[3], Fang-Hui Hou[5], Wu Chen[3]✉, Lihua Xiao[1,2]✉ & Yongyi Shen[1,2]✉

The outbreak of COVID-19 poses unprecedent challenges to global health[1]. The new coronavirus, SARS-CoV-2, shares high sequence identity to SARS-CoV and a bat coronavirus RaTG13[2]. While bats may be the reservoir host for various coronaviruses[3,4], whether SARS-CoV-2 has other hosts remains ambiguous. In this study, one coronavirus isolated from a Malayan pangolin showed 100%, 98.6%, 97.8% and 90.7% amino acid identity with SARS-CoV-2 in the E, M, N and S genes, respectively. In particular, the receptor-binding domain within the S protein of the Pangolin-CoV is virtually identical to that of SARS-CoV-2, with one noncritical amino acid difference. Results of comparative genomic analysis suggest that SARS-CoV-2 might have originated from the recombination of a Pangolin-CoV-like virus with a Bat-CoV-RaTG13-like virus. The Pangolin-CoV was detected in 17 of 25 Malayan pangolins analyzed. Infected pangolins showed clinical signs and histological changes, and circulating antibodies against Pangolin-CoV reacted with the S protein of SARS-CoV-2. The isolation of a coronavirus that is highly related to SARS-CoV-2 in pangolins suggests that they have the potential to act as the intermediate host of SARS-CoV-2. The newly identified coronavirus in the most-trafficked mammal could represent a future threat to public health if wildlife trade is not effectively controlled.

As coronaviruses (CoVs) are common in mammals and birds[5], we used the whole genome sequence of SARS-CoV-2 (WHCV; GenBank accession No. MN908947) in a Blast search of SARS-relate CoV (SARSr-CoV) sequences in available mammalian and avian viromic, metagenomic, and transcriptomic data. This led to the identification of 34 highly related contigs in a set of pangolin viral metagenomes (Extended Data Table 1). Therefore, we have focused our subsequent search of SARSr-CoV in pangolins.

We obtained the lung tissues from four Chinese pangolins (*Manis pentadactyla*) and 25 Malayan pangolins (*Manis javanica*) in a wildlife rescue center during March-August 2019, and analyzed them for SARSr-CoV using RT-PCR with primers targeting a conservative region of β CoVs. RNA from 17 of the 25 Malayan pangolins generated the expected PCR product, while RNA from the Chinese pangolins failed to amplify. The positive Malayan pangolins were all from the first transport. They were brought into the rescue center at the end of March, and gradually showed signs of respiratory disease, including shortness of breath, emaciation, lack of appetite, inactivity, and crying. Furthermore, 14 of the 17 pangolins that tested positive died within a time-interval of 1.5 month. Plasma samples of four PCR-positive and four PCR-negative Malayan pangolins were used in the detection of IgG and IgM antibodies against SARS-CoV-2 using a double-antigen sandwich ELISA. One of the PCR positive sample reacted strongly with an OD_{450} value of 2.17 (cutoff value = 0.11, Extended Data Table 2). The plasma remained positive at the dilution of 1:80, suggesting that the pangolin was naturally infected a SARS-CoV-2-like virus. The other three PCR-positive pangolins had no detectable antibodies against SARS-CoV-2. It is possible that they died during the acute stage of disease before the appearance of antibodies. Comparing with one β CoV-negative Malayan pangolin, histological examinations of tissues from four β CoV-positive Malayan pangolins revealed diffuse alveolar damage of various severity in the lung. In one case, alveoli were filled with desquamated epithelial cells and some macrophages with hemosiderin pigments, with significantly reduced alveolar space, leading to the consolidation of the lung. In other cases such changes were more focal (Fig 1 and Extended Data Fig 1). The severe case also had exudate with red blood cells and necrotic cell debris in bronchioles and bronchi. Focal mononuclear cell infiltration was seen in the bronchioles and bronchi of two of the cases, and hemorrhage was seen in bronchioles and small bronchi of one case (Extended Data Figs 1-3). Hyaline membrane and syncytia were not detected in alveoli of the four cases examined.

To isolate the virus, supernatant from homogenized lung tissue from one dead Malayan pangolin was inoculated into Vero E6 cells. Clear cytopathogenic effects were observed in cells after 72 hours incubation.

[1]College of Veterinary Medicine, South China Agricultural University, Guangzhou, 510642, China. [2]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, 510642, China. [3]Guangzhou Zoo, Guangzhou, 510070, China. [4]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China. [5]Guangdong Provincial Wildlife Rescue Center, Guangzhou, 510520, China. [6]Zhaoqing Branch Center of Guangdong Laboratory for Lingnan Modern Agricultural Science and Technology, Zhaoqing, 526238, China. [7]These authors contributed equally: Kangpeng Xiao, Junqiong Zhai. ✉e-mail: guangzhouchenwu@sina.com; lxiao@scau.edu.cn; shenyy@scau.edu.cn

Viral particles were detected by transmission electron microscopy mostly inside double-membrane vesicles, with a few outside them. They showed the typical coronavirus morphology (Fig 1e). RT-PCR targeting the spike (S) and RdRp genes produced the expected PCR products. The PCR products had ~84.5% and 92.2% nucleotide sequence identity to the partial S and RdRp genes of SARS-CoV-2, respectively.

Illumina RNAseq was used to identify viruses in the lung from nine pangolins. The mapping of sequence data to the reference SARS-CoV-2 (WHCV) genome identified CoV sequence reads in seven samples (Extended Data Table 3). For one samples, higher genome coverage was obtained by remapping the total reads to the reference genome (Extended Data Fig 4). We obtained the completed CoV genome (29,825 Bp), which was designated as Pangolin-CoV, using the assembled contigs, short sequence reads, and targeted PCR analysis. The full S gene in six PCR-positive samples were sequenced, revealing the presence of only four nucleotide differences in the sequence alignment (Extended Data Fig 5), indicating the presence of only one type of coronavirus in the batch of study samples. The predicted S, E, M and N genes of Pangolin-CoV are 3,798, 228, 669 and 1,260 bp in length, respectively, and share 90.7%, 100%, 98.6% and 97.8% amino acid identity to SARS-CoV-2 (Table 1).

In a Simplot analysis of whole genome sequences, Pangolin-CoV was highly similar to SARS-CoV-2 and Bat SARSr-CoV RaTG13, with sequence identity between 80 and 98%, except for the S gene (Fig 2). Further comparative analysis of the S gene sequences suggests that there were recombination events among some of the SARSr-CoV analyzed. In the region of nucleotides 1-914, Pangolin-CoV is more similar to Bat SARSr-CoV ZXC21 and Bat SARSr-CoV ZC45, while in the remaining part of the gene, Pangolin-CoV is more similar to SARS-CoV-2 and Bat-CoV-RaTG13 (Fig 2). In particular, the receptor-binding domain (RBD) of the S protein of Pangolin-CoV has only one amino acid difference from that of SARS-CoV-2. Overall, these data indicate that SARS-CoV-2 might have originated from the recombination of a Pangolin-CoV-like virus with a Bat-CoV-RaTG13-like virus (Fig 2). To further support this conclusion, we assessed the evolutionary relationships among β coronaviruses in the full genome, RdRp and S genes, and different regions of the S gene (Fig 2c and Extended Data Fig 6). The topologies mostly showed the clustering of Pangolin-CoV with SARS-CoV-2 and Bat SARSr-CoV RaTG13, with SARS-CoV-2 and Bat SARSr-CoV RaTG13 forming a subclade within the cluster (Fig 2c). In the phylogenetic analysis of the RBD, however, Pangolin-CoV and SARS-CoV-2 grouped together. Conflicts in cluster formation among phylogenetic analyses of different regions of the genome serve as a strong indication of genetic recombination, as previously seen in SARS-CoV and MERS-CoV[6,7].

As the S proteins of both SARS-CoV and SARS-CoV-2 have been shown to specifically recognize angiotensin converting enzyme II (ACE2) during the entry of host cells[2,8], we conducted molecular binding simulations of the interaction of the S proteins of the four closely related SARSr-CoVs with ACE2 proteins from humans, civets and pangolins. As expected, the RBD of SARS-CoV bands ACE2 from humans and civets efficiently in the molecular binding simulation. In addition, it appears to be capable of binding ACE2 of pangolins. In contrast, the S proteins of SARS-CoV-2 and Pangolin-CoV can potentially recognize ACE2 of only humans and pangolins (Extended Data Fig 7).

The SARS-CoV-2 is one of three zoonotic coronaviruses (the other two are SARS-CoV and MERS-CoV) infecting the lower respiratory tract and causing severe respiratory syndromes in humans[7,9]. It has been more contagious but less deadly than SARS-CoV thus far[10], with the total number of human infections far exceeding that caused by SARS-CoV[11]. Epidemiological investigations of the SARS-CoV-2 outbreak showed that some of the initial patients were associated with the Huanan Seafood Market, where live wildlife was also sold[10]. No animals thus far have been implicated as carriers of the virus. SARS-CoV-2 form clusters with SARS-CoV and bat SARS-related coronaviruses (Figure 2c). In addition, a bat coronavirus (Bat SARr-CoV RaTG13) has ~96% sequence identity to SARS-CoV-2 at the whole-genome level[2]. Therefore, it is reasonable to assume that bats are the native host of SARS-CoV-2, as previously suggested for SARS-CoVs and MERS-CoVs[12,13]. The SARSr-CoV virus identified in the present study and the metagenomic assemblies of viral sequences from Malayan pangolins[14] are genetically related to SARS-CoV-2, but are unlikely directly linked to the outbreak because of its substantial sequence differences from SARS-CoV-2. A virus related to Pangolin-CoV, however, appears having donated the RBD to SARS-CoV-2. SARSr-CoV sequences were previously detected in dead Malayan pangolins[15]. These sequences appear to be from Pangolin-CoV identified in the present study judged by their sequence similarity. In the present study, we have provided evidence on the potential for pangolins as the zoonotic reservoir of SARS-CoV-2-like coronaviruses. However, these pangolins showed clinical signs of disease. Generally, a natural reservoir host does not show severe disease, while an intermediate host may have clinical signs of infection[16]. Because of the lack of evidence from immunohistochemistry or in situ hybridization experiments, although a SARS-CoV-2-like coronavirus was detected in the lung of these pangolins, a direct association between the clinical signs or pathology and active virus replication is still not available. Experimental infection of healthy pangolins with pangolin-CoV would give us more definitive answers. However, pangolins are protected animals, making it difficult to carry out such experiments. Further studies are needed to confirm their roles in the transmission of SARSr-CoVs.

As the RBD of Pangolin-CoV is virtually identical to that of SARS-CoV-2, the virus in pangolins presents a potential future threat to public health. Pangolins and bats are both nocturnal animals, eat insects, and share overlapping ecological niches[17,18], which make pangolins the ideal intermediate host for some SARSr-CoV. Therefore, more systematic and long-term monitoring of SARSr-CoV in pangolins and other related animals should be implemented to identify the potential animal source of SARS-CoV-2 in the current outbreak.

Findings in the study support the call for stronger ban of illegal pangolin trade. Due to the insatiable demand for their meat as a delicacy and scales for use in traditional medicine in China, the illegal smuggling of other pangolins from Southeast Asia to China is rampant[18]. International co-operation and stricter regulations against illegal wildlife trade and consumption of game meat should be implemented. They can offer stronger protection of endangered animals as well as the prevention of major outbreaks caused by SARSr-CoVs.

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/s41586-020-2313-x.

1. Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* **382**, 727-733 (2020).
2. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
3. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* **17**, 181-192 (2019).
4. Banerjee, A., Kulcsar, K., Misra, V., Frieman, M. & Mossman, K. Bats and coronaviruses. *Viruses* **11**, 41 (2019).
5. Masters, P. S. The molecular biology of coronaviruses. *Adv Virus Res* **66**, 193-292 (2006).
6. 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).
7. Sabir, J. S. M. et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* **351**, 81-84 (2016).
8. Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454 (2003).
9. Song, Z. et al. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* **11**, E59 (2019).
10. Li, Q. et al. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *N Engl J Med* **382**, 1199-1207 (2020).

11. WHO. Coronavirus disease 2019 (COVID-19) Situation Report – 79. April 8, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200408-sitrep-79-covid-19.pdf?sfvrsn=4796b143_4.
12. Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-538 (2013).
13. Li, W. et al. Bats are natural reservoirs of SARS-Like coronaviruses. *Science* **310**, 676-679 (2005).
14. Lam, T. T.-Y. et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* (2020).
15. Liu, P., Chen, W. & Chen, J. P. Viral metagenomics revealed sendai virus and coronavirus infection of malayan pangolins (*Manis javanica*). *Viruses* **11**, 979 (2019).
16. Wu, D. et al. Civets are equally susceptible to experimental infection by two different severe acute respiratory syndrome coronavirus isolates. *J Virol* **79**, 2620-2625 (2005).
17. Nowak, R. M. *Walker's bats of the world*. (Johns Hopkins University Press, 1994).
18. Pantel, S. & Chin, S. Y. *Proceedings of the workshop on trade and conservation of pangolins native to South and Southeast Asia: 30 June-2 July 2008, Singapore Zoo*. (TRAFFIC Southeast Asia, 2009).

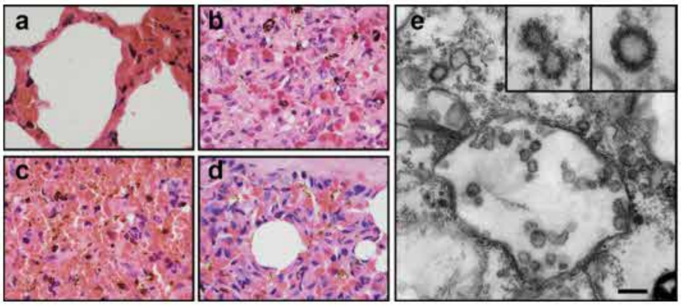


Fig. 1 | Pathological changes in the lungs of pangolins potentially induced by Pangolin-CoV. Histological changes in the lung tissues are compared between a negative Malayan pangolin (a) and three Malayan pangolins naturally infected with Pangolin-CoV (b-d, original magnification ×1000). Proliferation and desquamation of alveolar epithelial cells and hemosiderin pigments are seen in tissues from all three infected pangolins and severe capillary congestion is seen in one of them (c). Viral particles are seen in double-membrane vesicles in the transmission electron microscopy image (bar = 200 nm) taken from Verno E6 cell culture inoculated with supernatant of homogenized lung tissue from one pangolin (e), with morphology indicative of coronavirus (inserts at the upper right corner of e).

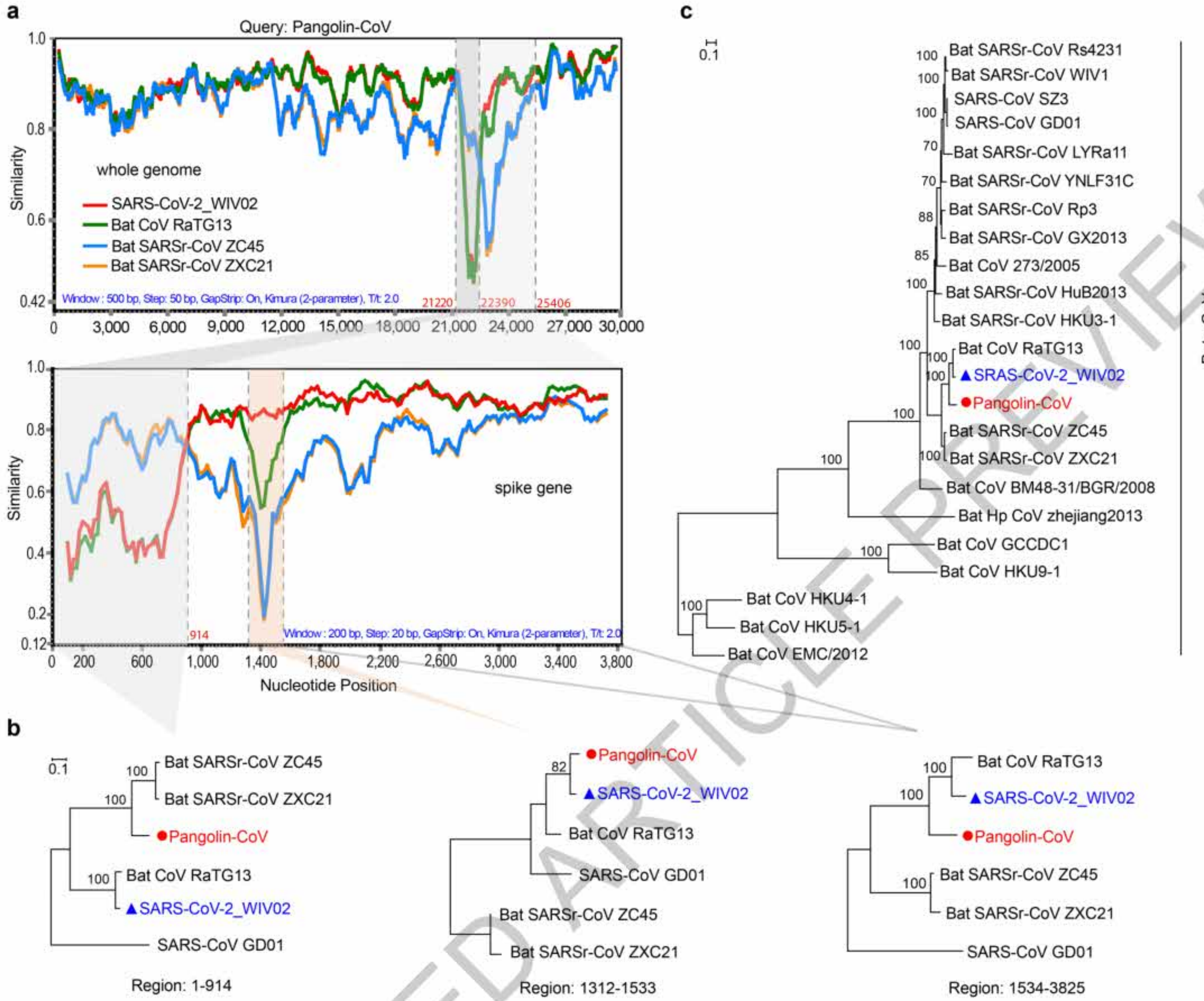


Fig. 2 | Genome characterization of Pangolin-CoV. (a) Similarity plot of the full-length genomes and S gene sequences of Pangolin-CoV against sequences of SARS-CoV-2_WIV02, Bat-CoV-RaTG13, Bat-CoV ZC45 and Bat-CoV ZXC21. While Pangolin-CoV has a high sequence identity to SARS-CoV-2 and Bat-CoV-RaTG13 in most regions of the S gene, it is more similar to Bat SARSr-CoV ZXC21 and Bat SARSr-CoV ZC45 at the 5′ end. (b) Because of the presence of genetic recombination, there is discrepancy in cluster formation among the outcomes of phylogenetic analyses of different regions of the S gene. (c) Phylogeny of coronaviruses closely related to SARS-CoV-2 based on full genome sequences. The phylogenetic tree was constructed using RAxML with the substitution model GTRGAMMAI and 1,000 bootstrap replicates. Numbers (>70) above or below branches are percentage bootstrap values for the associated nodes. The scale bar represents number of substitutions per site. Red circles indicate the pangolin coronavirus sequences generated in this study, while blue triangles indicate SARS-CoV-2 sequences from humans.

Table 1 | Genomic comparison of Pangolin-CoV with SARS-CoV-2, SARS-CoVs and Bat SARSr-CoVs (nt/aa %)

	S	E	M	N	Full-length genome
SARS-CoV-2 WHCV	84.5/90.7	99.1/100	93.2/98.6	96.1/97.8	90.1
SARS-CoV GD01	72.2/77.2	93.5/93.5	85.8/90.0	87.5/90.0	81.6
Bat SARSr-CoV RaTG13	88.5/89.8	99.6/100	93.6/99.1	94.0/96.7	88.9
Bat SARSr-CoV ZC45	83.1/86.1	98.7/100	94.2/99.6	88.9/93.3	88.0
Bat SARSr-CoV ZXC21	81.1/85.4	98.7/100	94.2/99.6	88.9/93.3	88.4

Methods

Metagenomic analysis and viral genome assembly

We collected viromic, metagenomic, and transcriptomic data of different mammals and birds in public databases, including NCBI Sequence Read Archive (SRA) and European Nucleotide Archive (ENA), for searching potential coronavirus sequences. The raw reads from the public databases and some inhouse metagenomic datasets were trimmed using fastp (v0.19.7)[19] to remove adaptor and low-quality sequences. The clean reads were mapped to the SARS-CoV-2 reference sequence (MN908947) using BWA-MEM (v0.7.17)[20] with > 30% matches. The mapped reads were harvested for downstream analyses. Contigs were *de novo* assembled using Megahit (v1.0.3)[21] and identified as SARS-CoV-2-related using BLASTn with E-values < 1e-5 and sequence identity > 90%.

Samples

Pangolins used in the study were confiscated by Customs and Department of Forestry of Guangdong Province in March and August 2019. They included four Chinese pangolins (*Manis pentadactyla*) and 25 Malayan pangolins (*Manis javanica*). The first transport confiscated contained 21 Malayan pangolins, while the second transport contained 4 Malayan pangolins and 4 Chinese pangolins. These animals were sent to the wildlife rescue center, and were mostly inactive and sobbing, and eventually died in custody despite exhausting rescue efforts. Tissue samples were taken from the lung of pangolins that had just died for histological and virological examinations.

Pathological examinations

Histological examinations were performed on lung tissues from five Malayan pangolins. Briefly, the tissues collected were cut into small pieces and fixed in 10% buffered formalin for 24 hrs. They were washed free of formalin, dehydrated in ascending grades of ethanol, and cleared with chloroform, and embedded with molten paraffin wax in a template. The tissue blocks were sectioned with a microtome. The sections were transferred onto grease-free glass slides, deparaffinized, and rehydrated through descending grades of ethanol and distilled water. They were stained with a hematoxylin and eosin staining kit (Baso Diagnostics Inc., Wuhan Servicebio Technology Co., Ltd.). Finally, the stained slides were mounted with coverslips and examined under an Olympus BX53 equipped with an Olympus PM-C 35 camera.

Virus isolation and RT-PCR analysis

Lung tissue extract from pangolins was inoculated into Vero E6 cells for virus isolation. The cell line was tested free of mycoplasma contamination using LookOut Mycoplasma PCR Detection Kit (SIGMA), and was authenticated by microscopic morphologic evaluation. Cultured cell monolayers were maintained in Dulbecco's Modified Eagle Medium (DMEM)/Ham's F-12. The inoculum was prepared by grounding the lung tissue in liquid nitrogen, diluting it 1:2 with DMEM, filtered through a 0.45 μm filter (Merck Millipore), and treated with 16 μg/ml trypsin solution. After incubation at 37°C for 1 hour, the inoculum was removed from the culture and replaced with fresh culture medium. The cells were incubated at 37 °C and observed daily for cytopathic effects.

Viral RNA was extracted from the lung tissue using the QIAamp® Viral RNA Mini kit (Qiagen) following the manufacturer-recommended procedures, and examined for CoV by RT-PCR using a pair of primers (F: 5′-TGGCWTATAGGTTYAATGGYATTGGAG-3′, R: 5′-CCGTCGATT GTGTGWATTTGSACAT-3′) designed to amplify the S gene of β CoV.

Transmission electron microscopy

Cell cultures that showed cytopathic effects were examined for the viral particles using transmission electron microscopy. Cells were harvested from the culture by centrifugation at 1,000× g for 10 min, and fixed initially with 2.5% glutaraldehyde solution at 4 °C for 4 hours, and again with 1% osmium tetroxide. They were dehydrated with graded ethanol and embedded with PON812 resin. Sections (80 nm in thickness) were cut from the resin block and stained with uranyl acetate and lead citrate sequentially. The negative stained grids and ultrathin sections were observed under a HT7800 transmission electron microscope (Hitachi).

Serological test

Plasma samples from eight Malayan pangolins were tested for anti-SARS-CoV-2 antibodies using a double-antigen ELISA kit for the detection of antibodies against SARS-CoV-2 by Hotgen (Beijing, China), following manufacturer-recommended procedures. The assay was designed for the detection of both IgG and IgM antibodies against SARS-CoV-2 in humans and animals, and marketed as supplemental diagnostic tool for COVID-19. It employs the capture of antibodies against SARS-CoV-2 by the S1 antigen precoated on ELISA plates, and the detection of the antibodies through the use of horseradish peroxidase-conjugated RBD. Both the S1 antigen and RBD fragment were expressed in eukaryotic cells. Data generated by the test developer have shown a 95% detection rate in the analysis of sera from over 200 COVID-19 patients. The assay has an inter-test variation of ≤15%, and no cross-reactivities with sera/plasma from patients positive for SARS-CoV, common and avian influenza viruses, mycoplasma, and chlamydia. Fifty microliters of plasma was analyzed in duplicate, together with two negative controls and one positive control. The reaction was read on a Synergy HTX Multi-Mode Microplate Reader (BioTek, USA) at 450/630 nm, with OD values being calculated. The cutoff OD value for positivity was 0.105 + mean OD from the negative controls, while the OD for the positive control should be ≥0.5. Positive samples were tested again with serial-diluted plasma.

Metagenomic sequencing

The lung tissue was homogenized by vortex with silica beads in 1 mL of phosphate-buffered saline. The homogenate was centrifuged at 10,000× g for 5 min, with the supernatant being filtered through a 0.45 μm filter (Merck Millipore) to remove large particles. The filtrate or virus culture supernatant was used in RNA extraction with the QIAamp® Viral RNA Mini kit. cDNA was synthesized from the extracted RNA using PrimeScriptScript II reverse transcriptase (Takara) and random primers, and amplified using Klenow Fragment (New England Biolabs). Sequencing libraries were prepared with NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs), and sequenced paired-end (150-bp) on an Illumina NovaSeq 6000. Specific PCR assays were used to fill genome sequence gaps, using primers designed based on sequences flanking the gap.

Phylogenetic analysis

Multiple sequence alignments of all sequence data were constructed using MAFFT v7.221[22]. The phylogenetic relationship of the viral sequences was assessed using RAxML v.8.0.14[23]. The best-fit evolutionary model for the sequences in each dataset was identified using ModelTest[24]. Potential recombination events and the location of possible breakpoints in β coronavirus genomes were detected using Simplot (version 3.5.1)[25] and RDP 4.99[26].

Molecular simulation of interactions between RBD and ACE2

The interaction between the RBD of the S protein of SARSr-CoV and ACE2 of humans, civets, and pangolins was examined using molecular dynamic simulation. The crystal structure of SARS-CoV RBD domain binding to human ACE2 protein complex was download from Protein Data Bank (PDB entry: 2AJF[27]). The structures of the complexes formed by ACE2 of civets or pangolins and RBD of SARS-CoV-2, Bat-CoV-RaTG13, and Pangolin-CoV were made using the MODELLER program[28], and superimposed with the template (PDB: 2AJF). The sequence identity of SARS-CoV RBD (PDB: 6ACD) to RBD of SARS-CoV-2, Bat-CoV and Pangolin-CoV was 76.5%, 76.8% and 74.2%, respectively, while the sequence of ACE2 protein of humans to that of pangolins and civets was 85.4% and 86.9%, respectively.

The molecular dynamic simulations on RBD-ACE2 complexes were carried out using the AMBER 18 suite[29] and ff14SB force field[30]. After two stage minimization, NVT and NPT-MD, a 30-ns production MD simulation was applied, with the time step being set to 2fs and coordinate trajectories being saved every 3ps. The MM-GBSA[31] approach was used to calculate the binding free energy of each ACE2 protein to the RBD of S protein, using the python script MMPBSA.py[32] in the build-in procedure of AMBER 18 suite. The last 300 frames of all simulations were extracted to calculate the binding free energy that excludes the contributions of disulfide bond.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence reads generated in this study are available in the NCBI Sequence Read Archive (SRA) database under the BioProject accession PRJNA607174. The complete genome sequence of pangolin-CoV has been deposited in GISAID with the accession numbers EPI_ISL_410721.

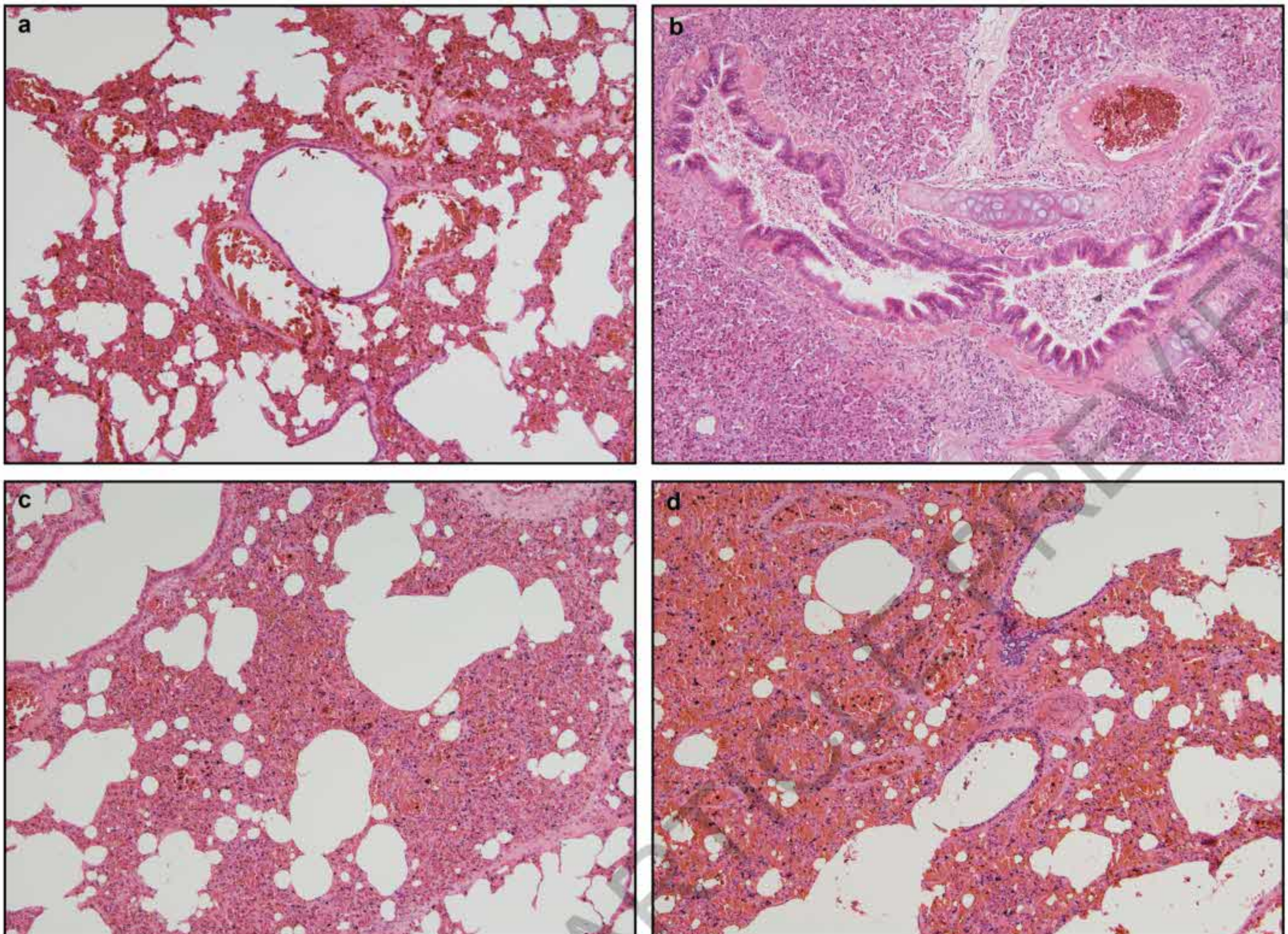
19. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
20. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
21. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).
22. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**, 1899-1900 (2010).
23. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).
24. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818 (1998).
25. Lole, K. S. et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152-160 (1999).
26. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol* **1**, vev003 (2015).
27. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864-1868 (2005).
28. Webb, B. & Sali, A. Comparative protein structure modeling Using MODELLER. *Curr Protoc Bioinformatics* **47**, 5.6.1-5.6.32 (2014).
29. Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. explicit solvent particle mesh ewald. *J Chem Theory Comput* **9**, 3878-3888 (2013).
30. Maier, J. A. et al. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-3713 (2015).
31. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* **10**, 449-461 (2015).
32. Miller, B. R., 3rd et al. MMPBSA.py: An efficient program for end-state free energy calculations. *J Chem Theory Comput* **8**, 3314-3321 (2012).

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 31822056 & 31820103014), National Key R&D Program of China (2017YFD0500404), Fund for the Key Program and Creative Research Group of the Department of Education of Guangdong Province (2019KZDXM004 and 2019KCXTD001), Guangdong Science and Technology Innovation Leading Talent Program (2019TX05N098), the Major Program of Guangdong Basic and Applied Research, the 111 Project (D20008), Ministry of Science and Technology of China, Chinese Academy of Engineering, Chinese Academy of Sciences, and Department of Agriculture of Guangdong Province.

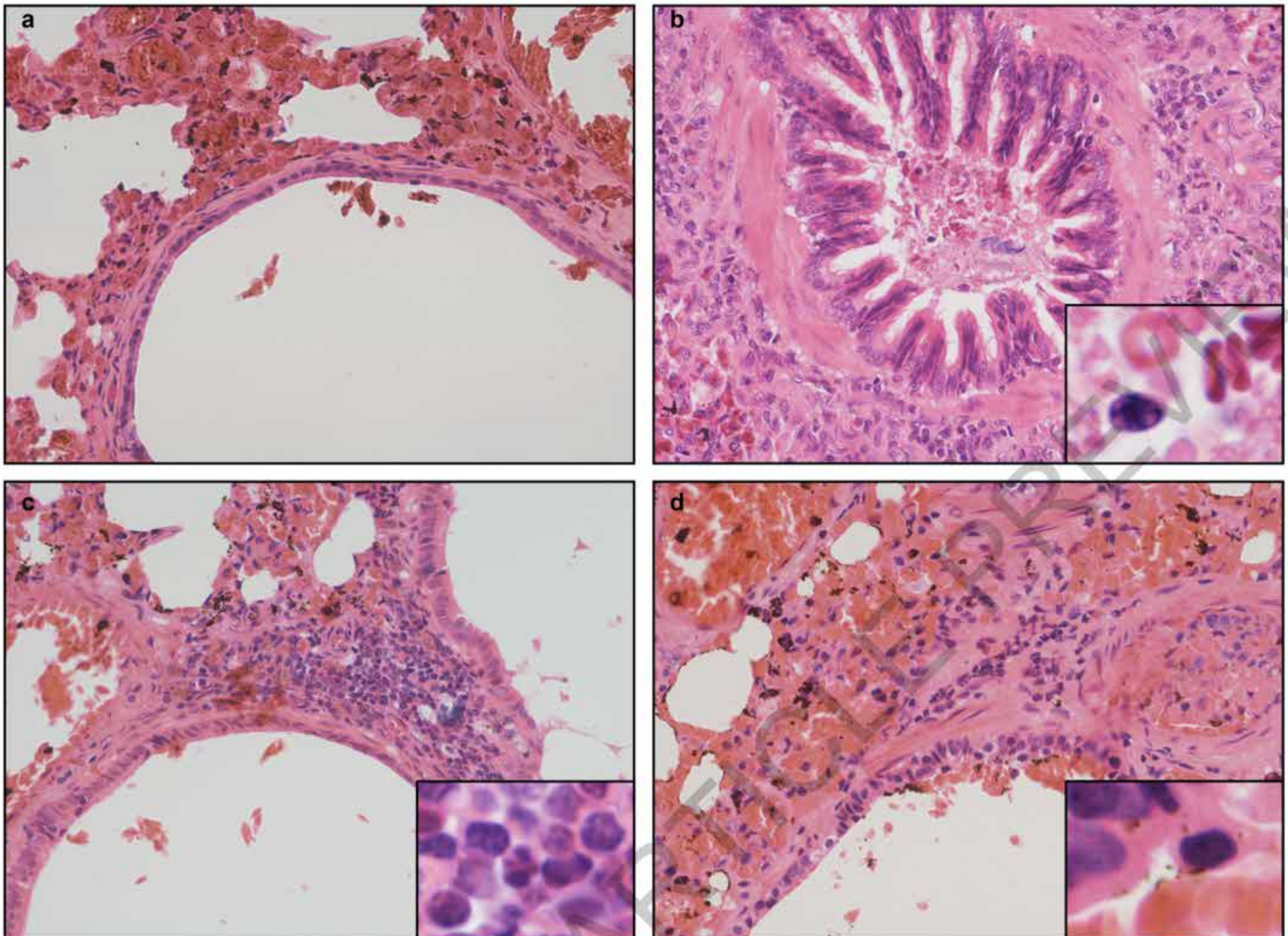
Author contributions YS, LX, and WC conceived the study; JJZ, FHH, YJW, SMP, MH, WJX, QHC and WC collected samples; JZ, NZ, XZ, NL, YG, XL, XS, ZZ, FS and WH performed virus isolation and sequencing; KX, YF, YL, ZiZ, and YS contributed to the analysis; YS and LX wrote the manuscript; YF and RAC edited the manuscript.

Competing interests The authors declare no competing interests.

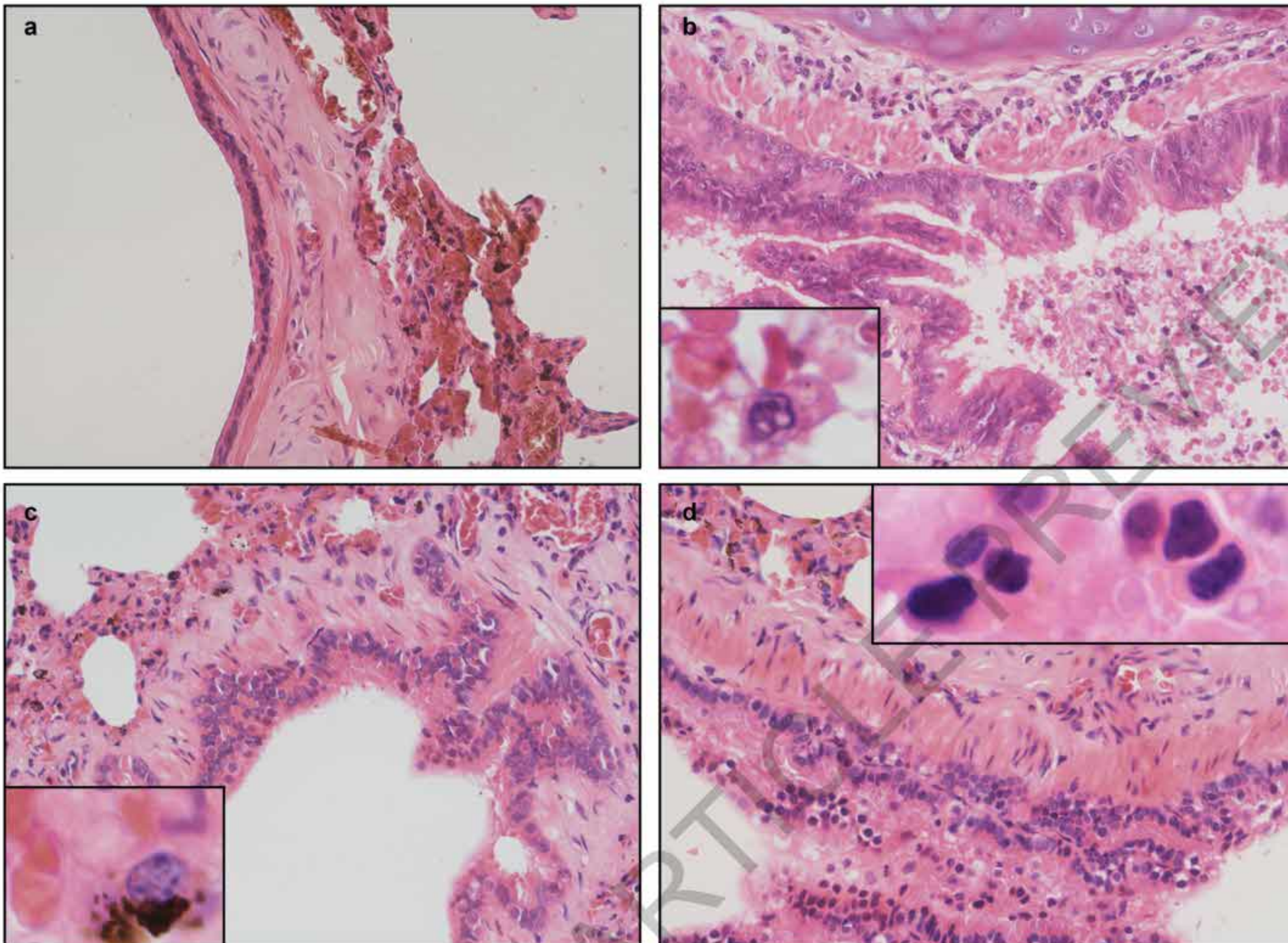
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2313-x.
Correspondence and requests for materials should be addressed to W.C., L.X. and Y.S.
Peer review information *Nature* thanks Andrew Ward and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



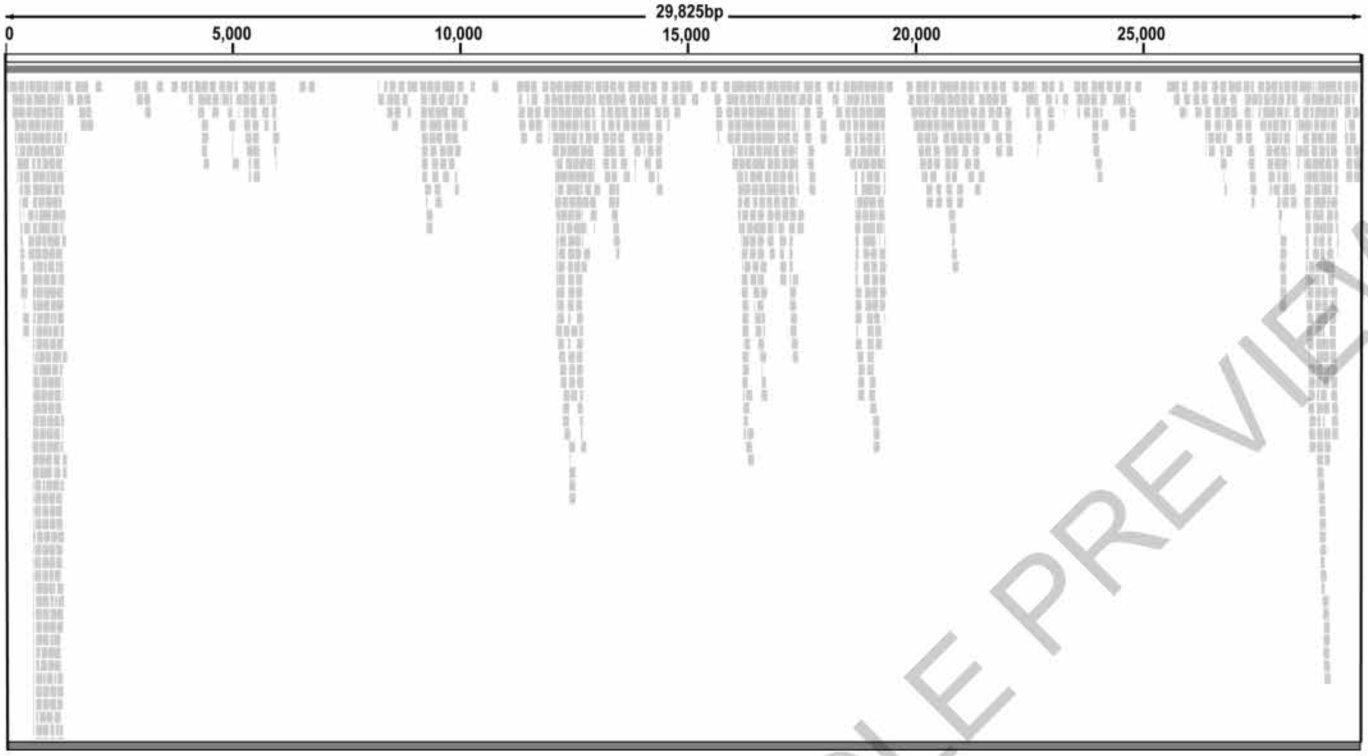
Extended Data Fig. 1 | Pathological changes in the lungs of pangolins. Three Malayan pangolins naturally infected with Pangolin-CoV (b-d, original amplification ×100) in comparison with the lung from a negative Malayan pangolin (a). Different degrees of consolidation are seen in the lung tissues from three infected pangolins (b-d), exudate is seen in the bronchi of one infected animal (b), and severe congestion is seen in the lung of one animal (d).



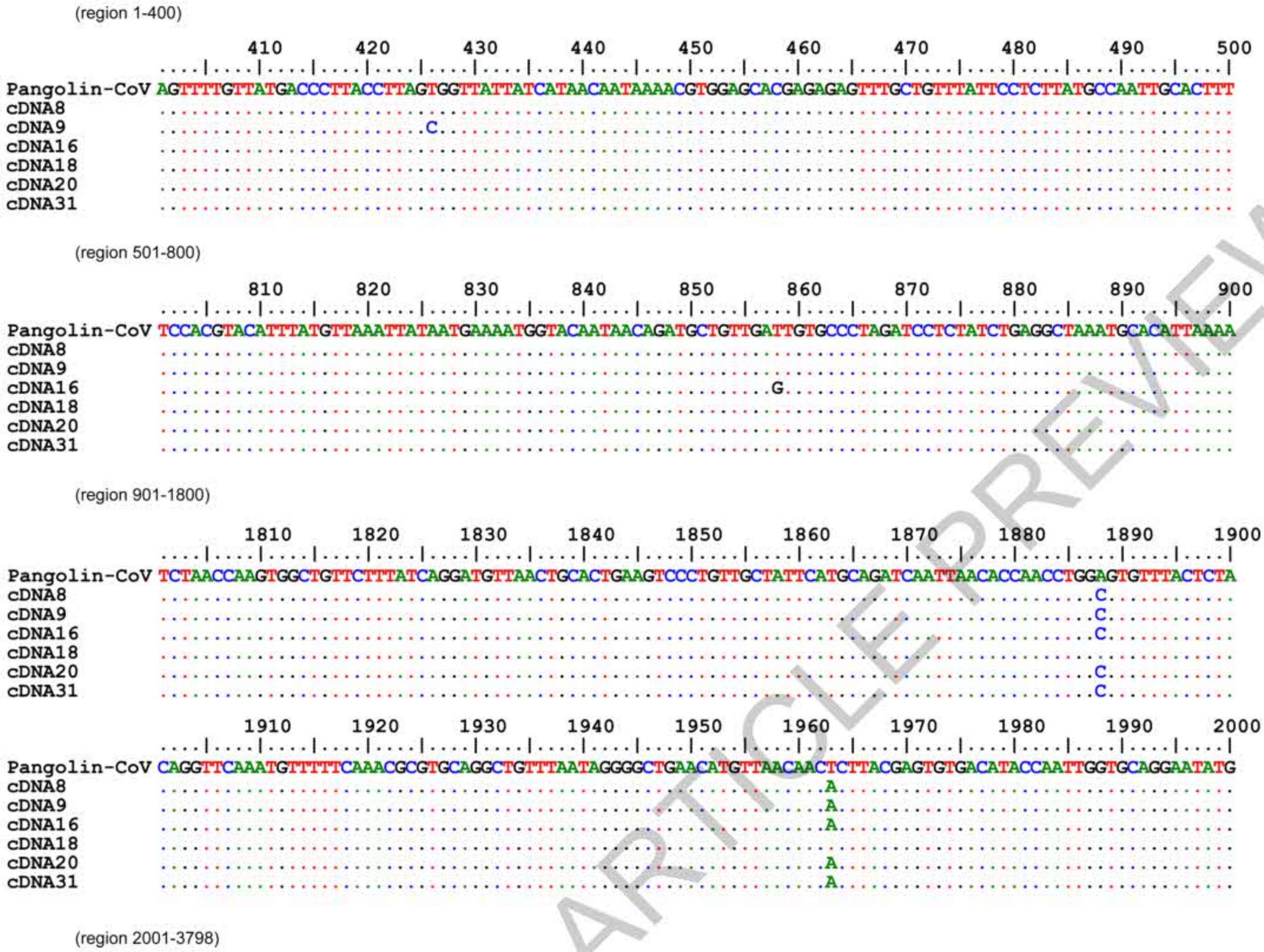
Extended Data Fig. 2 | Pathological changes in the bronchiole of pangolins. Three Malayan pangolins positive for Pangolin-CoV (b-d, original amplification ×100) in comparison with that from a negative Malayan pangolin (a). Red blood cells are seen in the bronchioles of two infected animals (b and d), mononuclear cell infiltration is seen in the bronchiole wall of one infected animal (c), and severe congestion is seen in the alveolar tissue (in close proximity of the bronchiole) of one animal (d). The respiratory epithelium in the bronchioles is still intact.



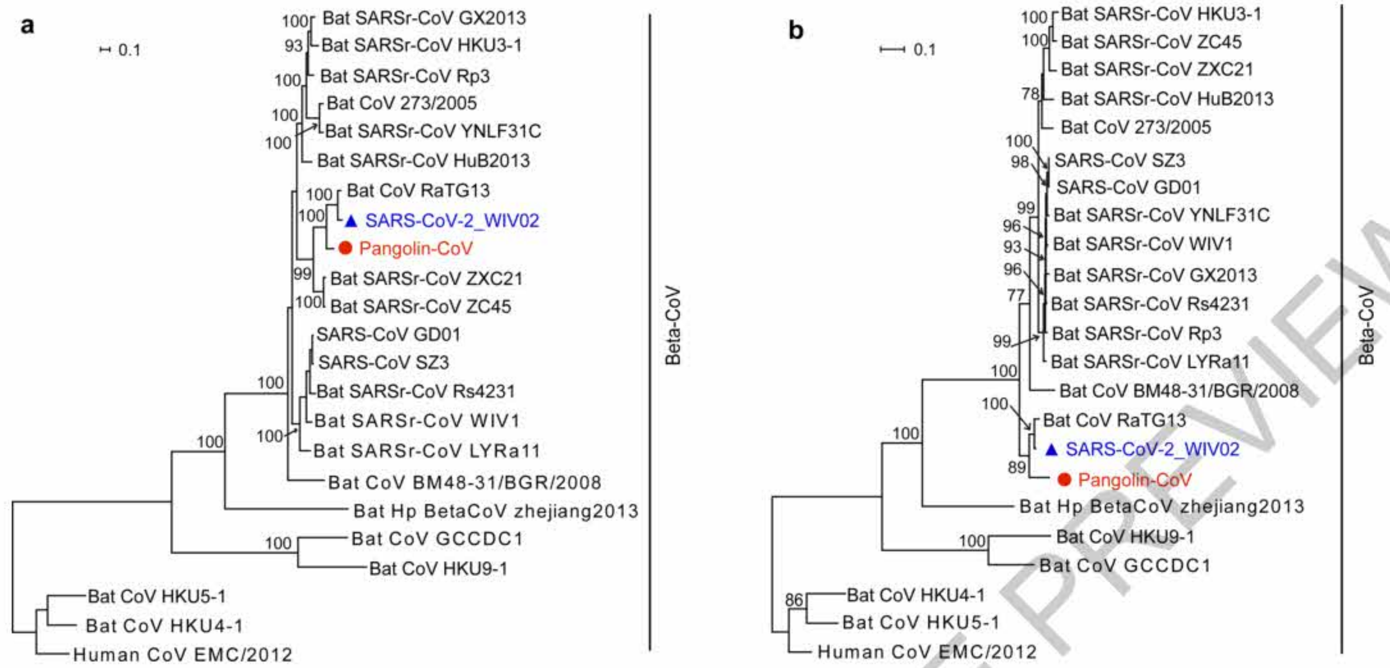
Extended Data Fig. 3 | Pathological changes in the bronchus of pangolins. Three Malayan pangolins positive for Pangolin-CoV (b-d, original amplification ×100) in comparison with that from a negative Malayan pangolin (a). Exudate with red blood cells is seen in the bronchus of one infected animal (b), macrophages with hemosiderin pigments and mononuclear cell infiltration are seen in the bronchus wall of two infected animals (c and d, respectively). The respiratory epithelium in the bronchi is still intact.



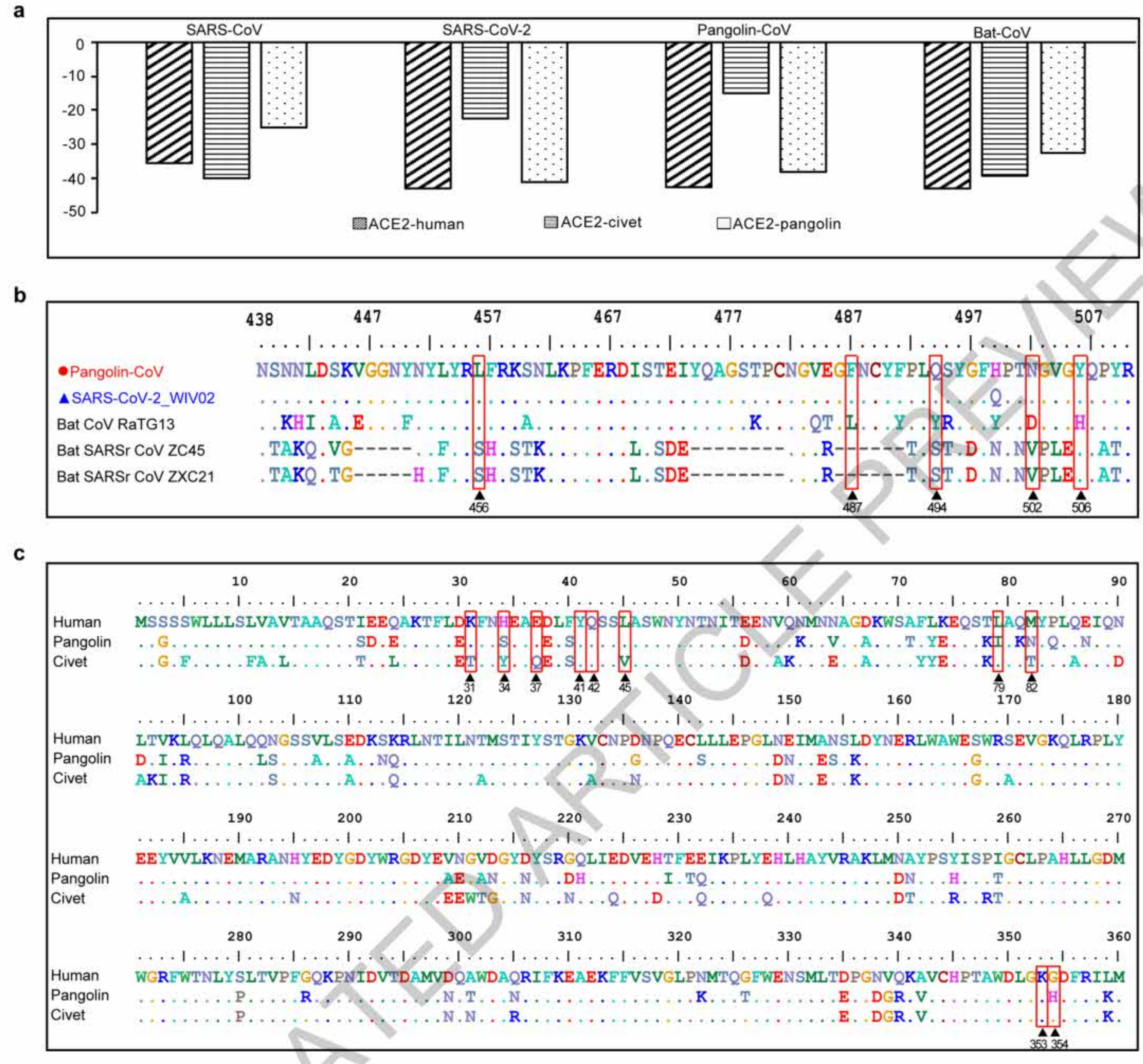
Extended Data Fig. 4 | Results of the mapping of raw reads from the high throughput sequencing of the pangolin lung tissue to the assembled Pangolin-CoV genome.



Extended Data Fig. 5 | Sequence polymorphism among nucleotide sequences of the full S gene among six positive lung samples from pangolins. Dots denote nucleotide identity to the reference sequence.



Extended Data Fig. 6 | Phylogeny of coronaviruses closely related to SARS-CoV-2. (a) Based on nucleotide sequences of the S gene; (b) Based on RdRp genes. The phylogenetic trees were constructed by RAxML with the substitution model GTRGAMMAI and 1,000 bootstrap replicates. Numbers (>70) above or below branches are percentage bootstrap values for the associated nodes. The scale bar represents the number of substitutions per site. Red circles indicate the pangolin coronavirus sequences generated in this study, while blue triangles indicate SARS-CoV-2 sequences from humans.



Extended Data Fig. 7 | Molecular binding simulations of the interaction of the S proteins of four closely related SARSr-CoVs with ACE2 proteins from humans, civets and pangolins. (a) Free energy (kcal/mol) for the binding of the RBD of S proteins of four SARSr-CoVs to ACE2 of potential hosts; (b) Alignment of the RBD sequences (key amino acids involved in interactions with ACE2 are boxed) of the S proteins from several genetically related SARSr-CoVs; (c) Alignment of partial ACE2 amino acid sequences (key amino acids involved in interactions with RBD are marked with arrowheads) from humans, pangolins and civets at their interface with the RBD of S proteins.

Extended Data Table 1 | Results of Blast search of SARSr-CoV sequences in available mammalian and avian viromic, metagenomic, and transcriptomic data using the SARS-CoV-2 sequence (GenBank accession No. MN908947)

Contig name	Sequence identity (%)	Length (bp)	E-value	Data source
con_9	98.623	363	0	PRJNA573298
con_15	97.303	519	0	PRJNA573298
con_2	96.97	203	2.41E-44	PRJNA573298
con_19	96.855	477	0	PRJNA573298
con_10	96.562	349	4.64E-168	PRJNA573298
con_1	95.96	203	1.12E-42	PRJNA573298
con_13	95.918	373	7.65E-42	PRJNA573298
con_11	95.804	402	5.63E-133	PRJNA573298
con_4	94.737	231	1.00E-38	PRJNA573298
con_25	94.737	779	0	PRJNA573298
con_22	94.297	655	5.73E-115	PRJNA573298
con_7	94.098	305	1.18E-133	PRJNA573298
con_16	94.041	387	3.08E-170	PRJNA573298
con_21	93.971	680	0	PRJNA573298
con_6	93.96	304	9.13E-130	PRJNA573298
con_24	93.74	610	0	PRJNA573298
con_20	93.343	721	0	PRJNA573298
con_12	93.333	373	4.14E-114	PRJNA573298
con_30	93.321	1048	0	PRJNA573298
con_29	92.892	886	0	PRJNA573298
con_5	92.632	231	2.17E-35	PRJNA573298
con_31	92.495	635	0	PRJNA573298
con_23	92.354	669	0	PRJNA573298
con_32	91.942	1031	0	PRJNA573298
con_34	91.884	1687	0	PRJNA573298
con_27	91.844	846	0	PRJNA573298
con_8	91.776	304	1.99E-121	PRJNA573298
con_3	91.705	218	6.83E-85	PRJNA573298
con_14	91.436	410	5.55E-158	PRJNA573298
con_18	91.429	385	5.23E-153	PRJNA573298
con_33	91.358	1177	7.06E-27	PRJNA573298
con_17	91.02	491	0	PRJNA573298
con_26	90.921	740	0	PRJNA573298
con_28	90.814	840	0	PRJNA573298

Extended Data Table 2 | OD values (450/630 nm) of ELISA testing of SARS-CoV-2 antibodies in eight pangolin plasma samples

Repetition / Sample	Repetition 1	Repetition 2	Average
1	2.253	2.088	2.1705
2	0.014	0.012	0.013
3	0.013	0.012	0.0125
4	0.023	0.025	0.024
5	0.01	0.012	0.011
6	0.011	0.011	0.011
7	0.028	0.028	0.028
8	0.053	0.052	0.0525
Negative control	0.01	0.01	0.01

Extended Data Table 3 | Identification of SARSr-CoV sequence reads in metagenomes from the lung of pangolins using the SARS-CoV-2 sequence (GenBank accession No. MN908947) as the reference

Sample ID	Animal species	Total reads*	No. mapped
M1	Malayan pangolin	107,267,359	496
M2	Malayan pangolin	38,091,846	302
M3	Malayan pangolin	79,477,358	14
M4	Malayan pangolin	32,829,850	1,100
M5	Malayan pangolin	547,302,862	56
M6	Malayan pangolin	232,433,120	10
M8	Malayan pangolin	44,440,374	12
M10	Malayan pangolin	227,801,882	0
Z1	Chinese pangolin	444,573,526	0

nature research

Corresponding author(s): Yongyi Shen, Lihua Xiao, Wu Chen

Last updated by author(s): Apr 22, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	*Provide a description of all commercial, open source and custom code used to analyse the data in this study, specifying the version used OR state that no software was used.*

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data that support the findings of this study have been deposited in GISAID with the accession numbers EPI_ISL_410721. Raw data of RNAseq are available from the NCBI SRA under the study accession number PRJNA607174.

October 2018

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging