ARTICLE

Check for updates

Single cell atlas for 11 non-model mammals, reptiles and birds

Dongsheng Chen[1,23], Jian Sun[2,3,23], Jiacheng Zhu[1,4,23], Xiangning Ding[1,4,23], Tianming Lan[1,5,23], Xiran Wang[2,3,23], Weiying Wu[1], Zhihua Ou[1], Linnan Zhu[1], Peiwen Ding [1,4], Haoyu Wang[1,4], Lihua Luo[1,4], Rong Xiang[1,4], Xiaoling Wang[1,4], Jiaying Qiu[1,4], Shiyou Wang [1,4], Haimeng Li[1,4], Chaochao Chai[1,4], Langchao Liang[1,4], Fuyu An[6], Le Zhang[7], Lei Han[7], Yixin Zhu[1,4], Feiyue Wang[1], Yuting Yuan[1], Wendi Wu[1], Chengcheng Sun[1,4], Haorong Lu[8,9], Jihong Wu[10,11,12], Xinghuai Sun[10,11,12], Shenghai Zhang[10,11,12], Sunil Kumar Sahu [1], Ping Liu[1], Jun Xia[1], Lijing Zhang[1,4], Haixia Chen[1,4], Dongming Fang[1], Yuying Zeng[1,4], Yiquan Wu[13], Zehua Cui[2,3], Qian He[2,3], Sanjie Jiang[1], Xiaoyan Ma[14], Weimin Feng[1], Yan Xu[1], Fang Li[15], Zhongmin Liu[15], Lei Chen[16], Fang Chen[1], Xin Jin [1], Wei Qiu[17], Tianjiao Wang[18], Yang Li[18], Xiumei Xing[18], Huanming Yang[1,19], Yanchun Xu[7,20], Yan Hua[6✉], Yahong Liu[2,3✉], Huan Liu [1,4,21✉] & Xun Xu [1,22✉]

The availability of viral entry factors is a prerequisite for the cross-species transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Large-scale single-cell screening of animal cells could reveal the expression patterns of viral entry genes in different hosts. However, such exploration for SARS-CoV-2 remains limited. Here, we perform single-nucleus RNA sequencing for 11 non-model species, including pets (cat, dog, hamster, and lizard), livestock (goat and rabbit), poultry (duck and pigeon), and wildlife (pangolin, tiger, and deer), and investigated the co-expression of *ACE2* and *TMPRSS2*. Furthermore, cross-species analysis of the lung cell atlas of the studied mammals, reptiles, and birds reveals core developmental programs, critical connectomes, and conserved regulatory circuits among these evolutionarily distant species. Overall, our work provides a compendium of gene expression profiles for non-model animals, which could be employed to identify potential SARS-CoV-2 target cells and putative zoonotic reservoirs.

A full list of author affiliations appears at the end of the paper.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent for coronavirus disease 2019 (COVID-19), which continues to threaten millions of lives worldwide[1–4]. Infected individuals without apparent clinical symptoms may also transmit the viruses[5]. Most patients infected by SARS-CoV-2 displayed symptoms of fever, dry cough, headache, dyspnea, and pneumonia[5]. Before the emergence of SARS-CoV-2 in December 2019, six coronaviruses (CoVs) are able to infect humans. These include four epidemic CoVs causing mild respiratory symptoms in human (i.e., HCoV-NL63, HCoV-229E, HCoV-OC43, and HCoV-HKU1) and two CoVs related to animal-to-human spillover events, including SARS-CoV and Middle East respiratory syndrome CoV. These six CoVs all originate from bats or rodents[6–11]. Currently, SARS-CoV-2 shares highest sequence similarity with a bat β-CoVs (BatCoV-RaTG13)[12], indicating a probable bat origin. Another mammalian animal, pangolins, has also been suggested as a potential host for SARS-CoV-2[13–15]. It is also reported that SARS-CoV-2 was capable of infecting cats, dogs, etc.[16,17]. However, the precise intermediate animals involved in the bat-to-human transmission of SARS-CoV-2 remain controversial and require continuous investigation.

Angiotensin-converting enzyme 2 (ACE2) has been recognized as the receptor for the spike protein of SARS-CoV, SARS-CoV-2, and HCOV-NL63[12,18,19]. Because of the diverse types of receptors and potential hosts of different viruses, understanding the tissue tropism and host range of a novel virus remains challenging. Regarding the high biosafety level required for the operation of highly pathogenic live viruses such as SARS-CoV-2, experiments involving cellular or animal models are mostly restricted to a small number of qualified laboratories, which hinders the large-scale investigation for the preferential tissues and hosts for these pathogens.

Single-cell sequencing has been applied to construct the single-cell atlas for a wide variety of species[20–26]. Previous studies suggested that animal tissues show high heterogeneity in terms of cellular composition and gene expression profiles, and ACE2 is only expressed in a small proportion of specific cell populations[27], thus revealing the potential application of single-cell analysis in investigating SARS-CoV-2 tropism. In this study, we applied single-nucleus RNA sequencing to determine the potential target cells and hosts for SARS-CoV-2. A comprehensive single-cell atlas was constructed for 11 species, comprising ~300,000 cells derived from a wide variety of anatomical locations, thus representing the broadest single-cell atlas to date.

We first systematically screened for putative SARS-CoV-2 target cells (indicated by the co-expression patterns of SARS-CoV-2 entry receptor ACE2 and SARS-CoV-2 entry activator TMPRSS2) to assess the potential tissue tropism and host range for the virus. In addition to screening virus entry factors, single-cell atlases for multiple species have been utilized to identify highly conserved regulomes and connectomes within evolutionarily distant species[28–30]. However, despite encouraging progress in state-of-the-art single-cell sequencing in traditional evolutionary biology and neural science, single-cell studies on non-neural tissues over a broad spectrum of non-model species are lacking.

Here we show extensive cellular cross-talk mediated by ligands and receptors, as well as dynamic intracellular regulatory circuits of critical pulmonary cell types in mammals, reptiles, and birds. Our findings could help narrow down suspected animal hosts of newly emerging viruses and accelerate the identification of animal species involved in virus amplification and interspecies transmission. Furthermore, our study could be employed to explore fundamental cellular and molecular networks among evolutionarily distant species.

Results

Generation of single-cell atlas for cat, tiger, and pangolin. Although the intermediate animal host involved in the emergence of SARS-CoV-2 remains obscure, cats, tigers, and pangolin were found permissive for CoV infection[13,17,31,32]. To establish a comprehensive transcriptome atlas for these three susceptible species, we generated single-nucleus libraries using 10× Genomics for various tissues (cat lung, kidney, liver, heart, eyelid, esophagus, and rectum; tiger lung, kidney, liver, spleen, and heart; and pangolin lung, kidney, liver, spleen, heart, esophagus, stomach, duodenum, and large intestine) (Fig. 1a and Supplementary Data 1). In total, 34,173, 80,608, and 92,863 single-cell transcriptomes passing quality control (see "Methods") were obtained for the cat (Fig. 1b), tiger (Fig. 1c), and pangolin (Fig. 1d), respectively (Supplementary Data 1). Cell clustering analysis was performed using Seurat[33,34] and cell type annotation was conducted according to the expression of canonical cell type markers (Fig. 1e–m, Supplementary Figs. 1–4, and Supplementary Data 1). Unsupervised clustering analysis revealed 30, 25, and 30 major cell types for the cat, tiger, and pangolin, respectively (Fig. 1b–d). The heart atlas for the three species mainly consisted of endothelial cells, fibroblasts, cardiomyocytes, and macrophages (Fig. 1e, h, k and Supplementary Figs. 1–3). The liver atlas for the three species primarily contained hepatocytes, Kupffer cells, hepatic stellate cells, endothelial cells, and liver sinusoidal endothelial cells (Fig. 1f, i and Supplementary Figs. 1–3). The spleen atlas for the tiger and pangolin included B cells, T cells, macrophages, and endothelial cells (Fig. 1j and Supplementary Figs. 2–3). The kidney atlas for the three species included proximal tubular cells, collecting duct cells, podocytes, Henle's loop cells, and endothelial cells (Fig. 1g and Supplementary Figs. 1–3). The cat and pangolin esophagus primarily contained endothelial, epithelial, and smooth muscle cells (Fig. 1m and Supplementary Figs. 1 and 3). The cat eyelid tissue consisted of Wolfring's gland, endothelial, epithelial, and immune cells (Supplementary Fig. 1), whereas the rectum tissue contained goblet, Paneth, enteroendocrine, and progenitor cells (Supplementary Fig. 1). In the pangolin digestive tract, we identified endothelial cells, epithelial cells, macrophages, secretory cells, and smooth muscle cells in the stomach (Supplementary Fig. 3); endothelial cells, enteroendocrine cells, goblet cells, Paneth cells, progenitor cells, stem cells, transit-amplifying (TA) cells, and tuft cells in the duodenum (Supplementary Fig. 3); and enterocyte cells, enteroendocrine cells, goblet cells, macrophages, progenitor cells, stem cells, and TA cells in the large intestine (Supplementary Fig. 3). Alveolar type 1 cells (AT1), alveolar type 2 cells (AT2), ciliated cells, secretory cells, endothelial cells, fibroblasts, T cells, B cells, and macrophages were identified in the lung atlas (Supplementary Fig. 4 and Supplementary Data 1). Taken holistically, we constructed the single-cell atlas for three important, yet poorly characterized, non-model species, which revealed their cellular taxonomies and thus laid the foundation for in-depth study regarding the cellular biology and development of these animals.

Investigation of SARS-CoV-2 tissue tropism. Although the susceptibility of cats, tigers, and pangolins to CoV has been reported[13,17,31,32], little information is available regarding the distribution of putative SARS-CoV-2 target cells in these animals. Utilizing the high-quality and comprehensive single-cell atlas for distinct tissues of the three species, we systematically evaluated the cellular distribution of SARS-CoV-2 entry factors.

In the cat, ACE2-expressing cells and TMPRSS2-expressing cells were observed in all the tissues sampled (Figs. 2a, c and 3a). ACE2 and TMPRSS2 co-expressing cells were detected in the kidney (collecting duct cells, endothelial cells, Henle's loop cells,
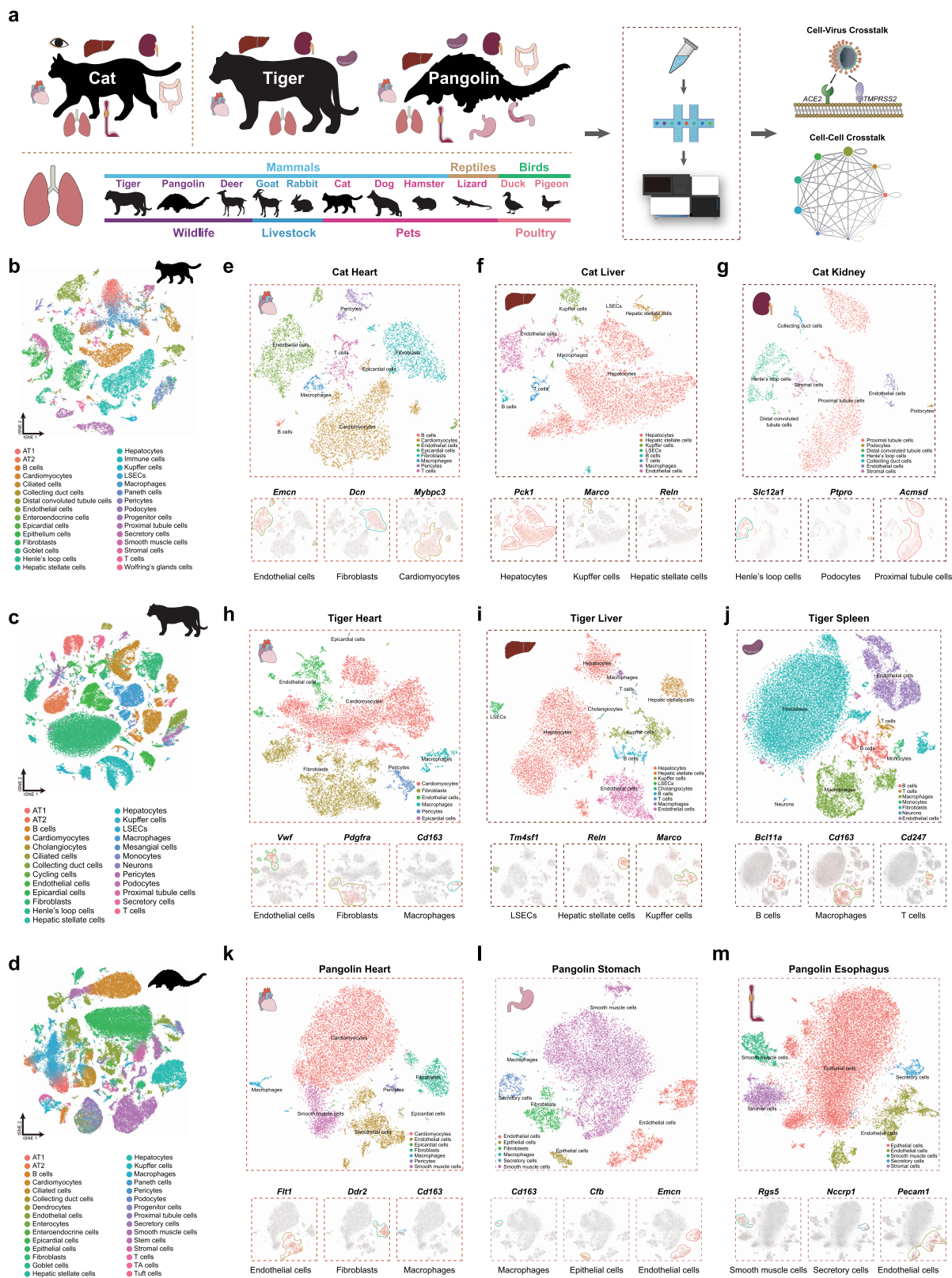
proximal tubule cells, and stromal cells), heart (epicardial cells), eyelid (endothelial cells, epithelial cells, and Wolfring's glands cells), rectum (enteroendocrine cells) (Fig. 2b, d), and lung (AT2 cells, ciliated cells, endothelial cells, fibroblasts, and secretory cells) (Fig. 3b–d). Notably, we observed the co-expression of *ACE2* and *TMPRSS2* in over 40% of kidney

proximal tubule cells (Fig. 2b and Supplementary Data 2). In addition, co-expression of *ACE2* and *TMPRSS2* was detected in more than 50% of Wolfring's gland cells in the cat eyelid and epicardial cells in the cat heart (Fig. 2b and Supplementary Data 2), suggesting the possibility that such tissue could act as a putative portal for the initial infection and transmission of

Fig. 1 Single-cell atlas for cat, tiger, and pangolin. a Illustration of the overall project design. Single-nucleus RNA-seq was performed with multi-tissues of cat (lung, kidney, liver, heart, eyelid, esophagus, and rectum), tiger (lung, kidney, liver, spleen, and heart), and pangolin (lung, kidney, liver, spleen, heart, esophagus, stomach, duodenum, and large intestine), with tissue types indicated by corresponding images. Lung tissues were used from other animals (deer, goat, rabbit, dog, hamster, lizard, duck, and pigeon). **b–d** tSNE plot showing single-cell atlas of cat (**b**), tiger (**c**), and pangolin (**d**). Dots with colors represent different cell types, which were indicated below. LSECs, liver sinusoidal endothelial cells. TA cells, Transit-amplifying cells. **e–g** Plot (upper panel) showing different cell types in cat heart, liver, and kidney. Feature plot (lower panel) showing expression of marker genes (above) of indicated cell types (below). **h–j** Plot (upper panel) showing different cell types in tiger heart, liver, and spleen. Feature plot (lower panel) showing expression of **h** marker genes (above) of indicated cell types (below). **k–m** Plot (upper panel) showing different cell types in pangolin heart, stomach, and esophagus. Feature plot (lower panel) showing expression of marker genes (above) of indicated cell types (below). In feature plot, red dots representing cells expressed the marker gene and corresponding cell types were highlighted with colored lines.

SARS-CoV-2. Thus, our data showed that cats may be prone to multiple organ (lung, kidney, heart, rectum, and eyelid) infection by SAR-CoV-2 when exposed.

In the tiger, *ACE2*-expressing cells and *TMPRSS2*-expressing cells were observed in all the tissues sampled (Fig. 2e, g). *ACE2* and *TMPRSS2* co-expression was detected in five out of eight kidney cell types (collecting duct cells, Henle's loop cells, mesangial cells, podocytes, and T cells), one out of six heart cell types (endothelial cells) (Fig. 2f, h and Supplementary Data 2), and six out of nine lung cell types (AT1 cells, AT2 cells, ciliated cells, endothelial cells, fibroblasts, and secretory cells) (Fig. 3b, e, f), thus indicating infection potential of the lung, kidney, and heart.

In the pangolin, *ACE2*-expressing cells and *TMPRSS2*-expressing cells were observed in all the tissues sampled (Fig. 2i, k). Potential SARS-CoV-2 target cells were mainly found in the kidney (collecting duct cells, endothelial cells, podocytes, and proximal tubule cells), liver (hepatocytes), spleen (macrophages), and stomach (secretory cells) (Fig. 2j, l and Supplementary Data 2). Unlike the tiger and cat, however, SARS-CoV-2 target cells were only present in a small proportion of pangolin lung endothelial cells, with other lung cell types absent of *ACE2* and *TMPRSS2* co-expression (Fig. 3b and Supplementary Data 2). Similar to the cat, exposure to SARS-CoV-2 may lead to systemic infection in the pangolin.

Collectively, we evaluated SARS-CoV-2 tissue tropism in the cat, tiger, and pangolin. Distributions of putative SARS-CoV-2 target cells were restricted in specific cell types of certain tissues, with skewed expression towards the respiratory and urinary systems in the cat, respiratory system in the tiger, and urinary system in the pangolin.

Screening of SARS-CoV-2 entry factors in lung atlases. Lung is one of the main target tissue of various respiratory viruses[35] and pneumonia is a typical symptom of COVID-19[36]. To characterize lung cell compositions in a broad range of species, we generated single-nucleus libraries of lung cells for 11 species covering livestock, poultry, pets, and wildlife (Fig. 3a and Supplementary Fig. 4), resulting in a total of 114,015 pulmonary cells passing quality control (Supplementary Data 1).

Recent comparative and structural analysis of ACE2 predicted a broad SARS-CoV-2 host range in vertebrates[37]. Consistently, functional and genetic analysis of ACE2 orthologs among mammals suggested a broad potential host tropism of SARS-CoV-2[38]. To evaluate the expression patterns of SARS-CoV-2 entry factors in lung cells for various species, we screened eight mammalian species (cat, tiger, pangolin, dog, hamster, deer, rabbit, and goat) (Fig. 3a and Supplementary Data 3). Results suggested that, in addition to the three previously mentioned species (cat, tiger, and pangolin) (Fig. 2), four other mammalian species (dog, hamster, deer, and goat) also demonstrated co-expression of *ACE2* and *TMPRSS2*[39] in specific cell types (Fig. 3b and Supplementary Data 3). Putative SARS-CoV-2 target cells

were detected in three cell types in the dog (AT1, AT2, and ciliated cells). In the hamster lung, putative SARS-CoV-2 target cells were abundant in ciliated cells and found in a small proportion of AT1 cells (Fig. 3g, h). Four deer cell types (AT1 cells, AT2 cells, fibroblasts, and endothelial cells) co-expressed *ACE2* and *TMPRSS2*. In the goat lung, *ACE2* and *TMPRSS2* co-expression was detected in AT1, ciliated, and endothelial cells (Fig. 3i, j). The proportions of SARS-CoV-2 target cells in the cat lung ciliated cells were higher than the proportions in corresponding cell type of the other species studied (Fig. 3b and Supplementary Data 21). Meanwhile, the proportion of cells co-expressing *ACE2* and *TMPRSS2* in the cat lung ciliated cells was higher than macrophage cells (Fig. 3b and Supplementary Data 22). Besides, we investigated the expression of SARS-CoV-2 entry factors in human lung[30] and detected co-expression of *ACE2* and *TMPRSS2* in AT2 and ciliated cells (Supplementary Fig. 5).

Next, we screened the expression patterns of orthologs of *ACE2* and *TMPRSS2* in the lung cells of two avian species (duck and pigeon) and one reptile species (lizard). Consistent with previous research that SARS-CoV-2 replicates poorly in ducks[40], no putative SARS-CoV-2 target cells were found in the lung cells of the poultry species (Fig. 3a, b). Intriguingly, *ACE2* and *TMPRSS2* co-expression was detected in the AT2 and ciliated cells of the lizard lung (Fig. 3a, b). As the binding affinity of ACE2 to SARS-CoV-2 has not been evaluated in lizards, this result should be interpreted with caution.

Cross-talk landscape of cat cells and viruses. To reveal the putative cellular tropism of feline viruses or viruses that may potentially infect cats at single-cell resolution, we investigated the expression patterns of 13 virus entry factors (*Nectin1*, *Ncam1*, *Anpep*, *Ngfr*, *Grm2*, *Tnfrsf4*, *Flvcr1*, *Cxcr4*, *Nectin2*, *F11r*, *Slc20a1*, *Tfrc*, and *Slc20a2*) (Supplementary Data 4). Overall, the 13 viral receptors could be classified into three clusters (Fig. 4a, b). Briefly, cluster A receptors (*Ncam1* and *Anpep*) demonstrated a highly specific target cell spectrum, while cluster C receptors (*Flvcr1*, *Cxcr4*, *Nectin1*, *Nectin2*, *F11r*, *Slc20a1*, *Tfrc*, and *Slc20a2*) showed pan-expression patterns (Fig. 4a, b). In contrast, cluster B receptors (*Ngfr*, *Grm2*, and *Tnfrsf4*) were expressed at very low levels in all 31 feline cell types included in this study.

Different receptors for the same virus show high variability in terms of expression abundance and spatial distribution. For example, rabies lyssavirus (RABV) was reported to infect the kidney, lung, and heart of cats[41,42]. So far, three well-known receptors for RABV have been reported, i.e., *Ncam1*, *Ngfr*, and *Grm2*[43]. Here, *Ncam1* was enriched in kidney stromal cells, lung ciliated cells, and heart cardiomyocytes and fibroblasts, whereas *Ngfr* and *Grm2* were weakly expressed, indicating the diverse cellular tropism of RABV in cats. Both *Cxcr4* and *Tnfrsf4* are receptors of feline immunodeficiency virus (FIV). Here, *Cxcr4* was expressed in T and B cells of the heart and liver, in line with previous study showing that FIV can infect T and B
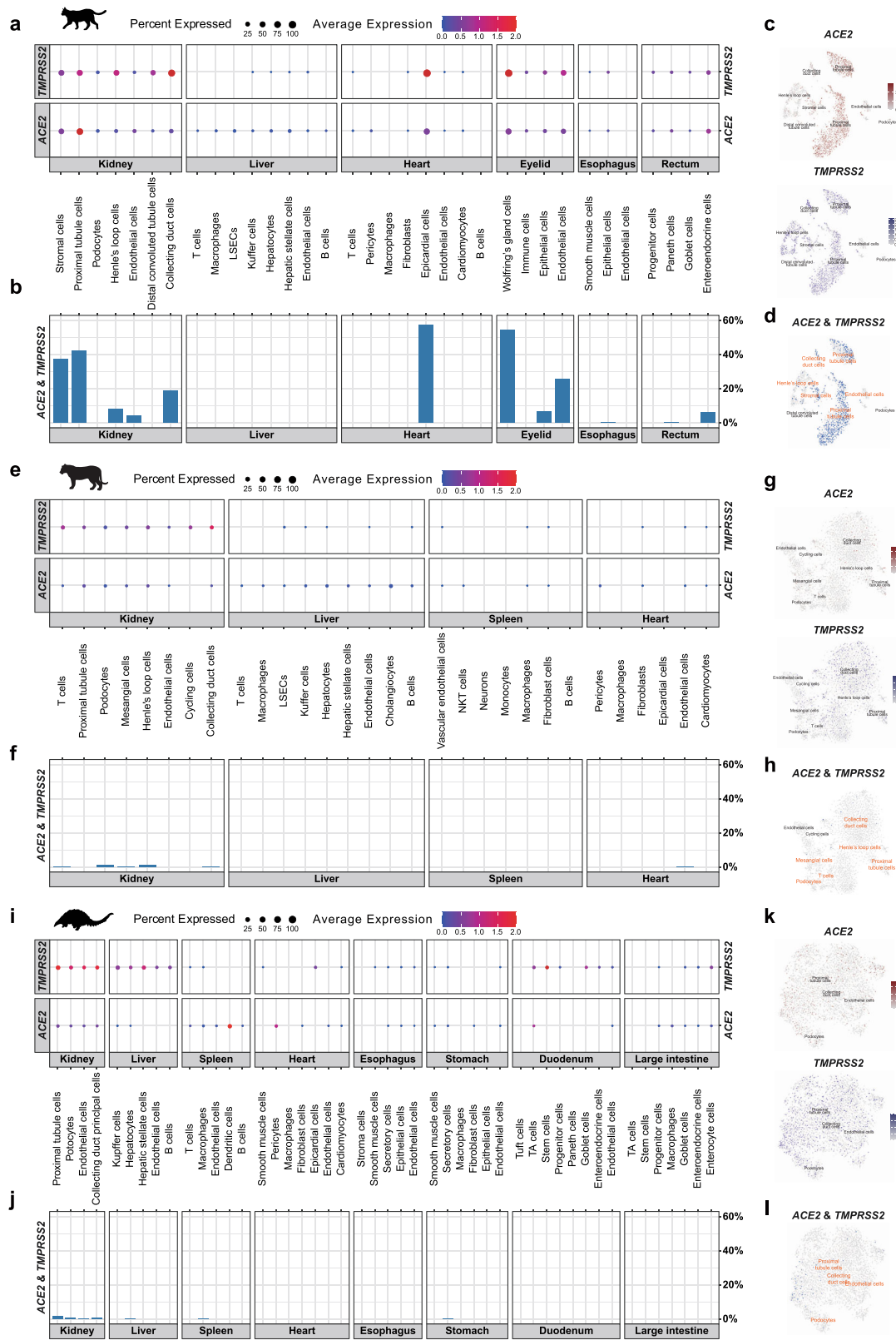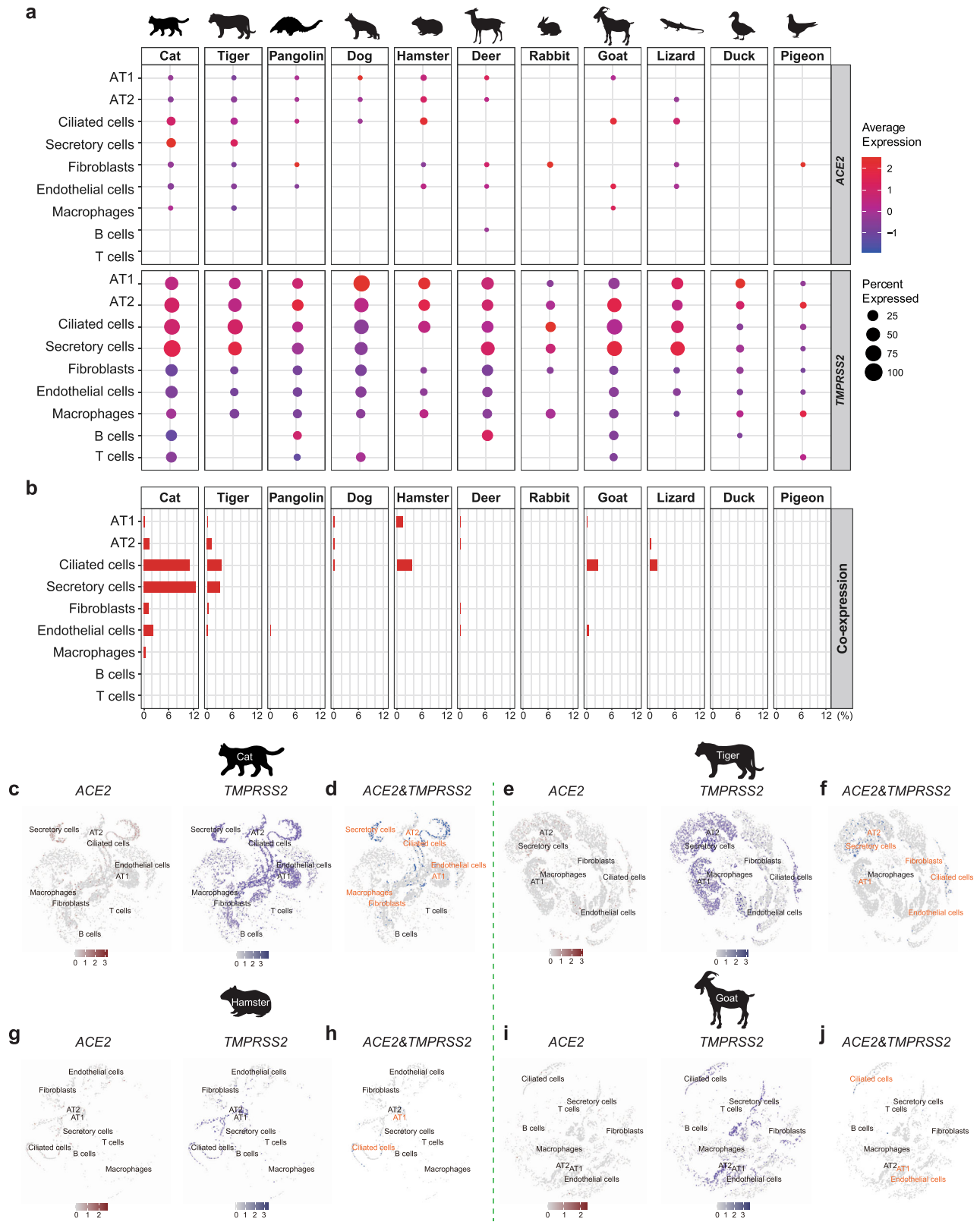
Fig. 2 Screening of putative SARS-CoV-2 target cells at single-cell resolution in various tissues of cat, tiger, and pangolin. a, e, i Dot plot showing the expression of *ACE2* and *TMPRSS2* in various cell types of cat (**a**), tiger (**e**), and pangolin (**i**). Each dot represents gene expression within indicated cell type, of which the color represents average normalized expression level and the size indicates the percentage of cells that expressed the gene within each cell type. **b, f, j** Bar plot showing percentage of cells that co-express *ACE2* and *TMPRSS2* within each cell type of cat (**b**), tiger (**f**), and pangolin (**j**). **c, g, k** Feature plot showing the expression pattern of *ACE2* and *TMPRSS2*, respectively, in indicated cell types in kidney of cat (**c**), tiger (**e**), and pangolin (**i**). **d, h, l** Feature plot showing the co-expression of *ACE2* and *TMPRSS2* in indicated cell types in kidney of cat (**d**), tiger (**h**), and pangolin (**l**). Putative SARS-CoV-2 target cells were highlighted with red-colored labels.
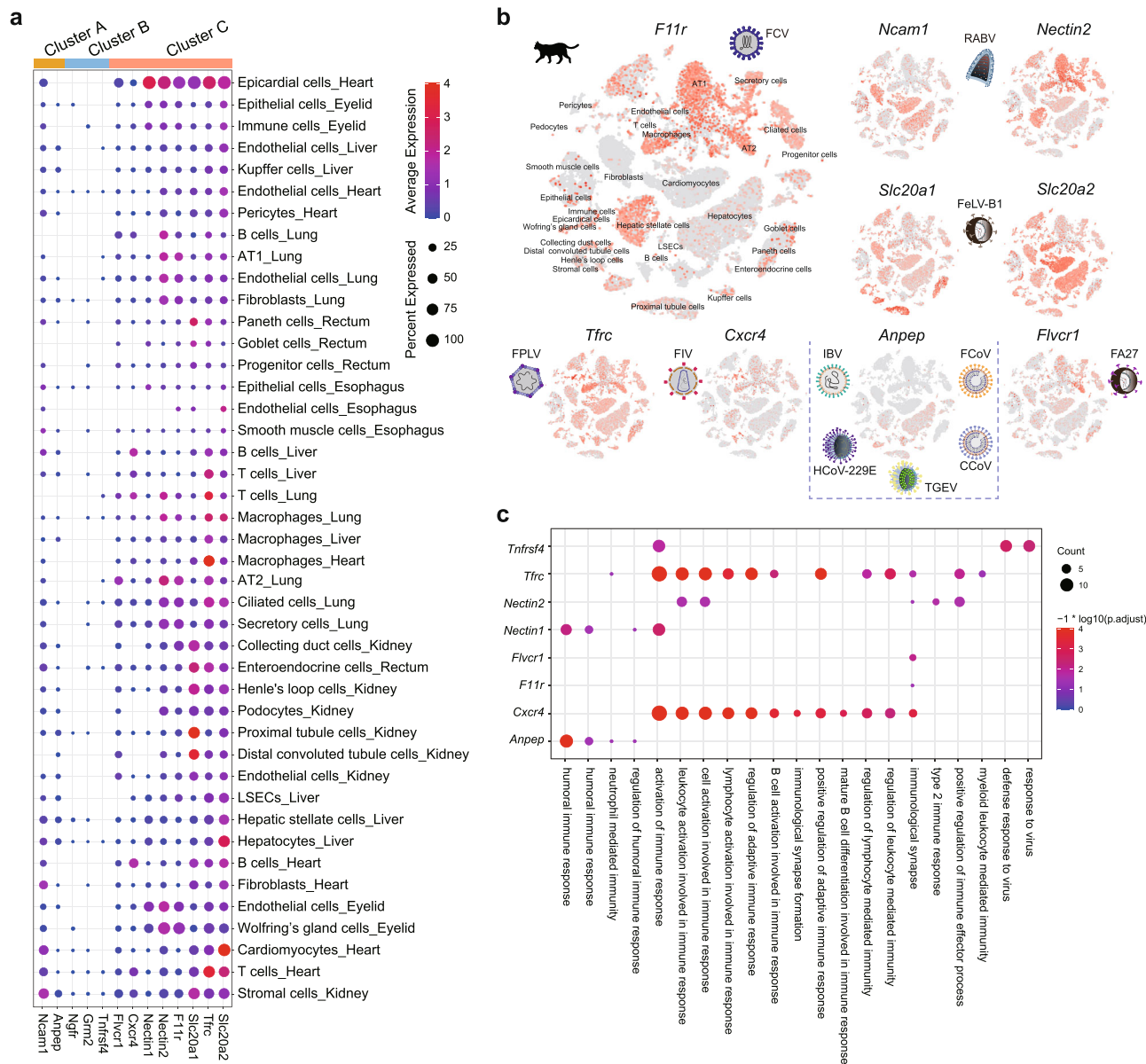
Fig. 3 Expression of *ACE2* and *TMPRSS2* in lung cells of wildlife, pets, livestock, and poultry. a Dot plot showing expression level and percentage of *ACE2* and *TMPRSS2* (indicated on the right) in each cell type (indicated on the left) of lungs across a variety of species (indicated above). Dot color represents scaled average expression level and dot size represent percentage of cells that expressed indicated gene within each cell type. **b** Bar plot showing percentage of cells that co-expressed *ACE2* and *TMPRSS2* within each cell type (indicated on the left) of different species (indicated above). **c**, **e**, **g**, **i** Feature plot showing the expression pattern of *ACE2* and *TMPRSS2*, respectively, in indicated cell types in lung of cat (**c**), tiger (**e**), hamster (**g**), and goat (**i**). **d**, **f**, **h**, **j** Feature plot showing the co-expression of *ACE2* and *TMPRSS2* in indicated cell types in lung of cat (**d**), tiger (**f**), hamster (**h**), and goat (**j**). Putative SARS-CoV-2 target cells were highlighted with red-colored labels.

Fig. 4 Cross-talk landscape of cat cells and representative feline viruses. a Dot plot showing expression percentages and average expression levels of virus receptors (indicated below) in each cell types (indicated on the right) of cat cells. **b** Feature plot showing the expression patterns of virus receptors (*F11r, Ncam1, Nectin2, Slc20a1, Slc20a2, Tfrc, Cxcr4, Anpep, Flvcr1*) in cat cells. Corresponding virus species and image were illustrated approximate to the feature plot. CCov, canine coronavirus; FA27, feline leukemia virus strain C; FCV, feline calicivirus; RABV, Rabies lyssavirus; FCoV, feline coronavirus; FeLV-B1, feline leukemia virus strain B; FPLV, feline panleukopenia virus; FIV, feline immunodeficiency virus; HCoV-229E, human coronavirus 229E; IBV, infectious bronchitis virus; TGEV, transmissible gastroenteritis virus. **c** Bubble plot showing the immune-related GO terms of eight representative virus receptors in cat. The dot color was positively correlated with enrichment levels, whereas dot size was determined by count of genes in that GO term. *P*-values were calculated using hypergeometric test. Multiple comparisons adjustment was performed using Benjamini and Hochberg method. Exact *P*-value and source data were included in the Source Data file.

lymphocytes[44]. In contrast, *Tnfrsf4* was expressed poorly in feline cells, implying a receptor usage preference of FIV.

Receptors for different strains of the same virus show a high degree of heterogeneity. For instance, *Slc20a1* and *Slc20a2* are the receptors for feline leukemia virus (FeLV) strain B/lambda-B1[45], whereas *Flvcr1* is the receptor for the FeLV strain C/FA27[46]. These three receptors were all present in the cluster C receptors (pan-expression cluster). Infection of FeLV in the myocardium, thymus, mesentery, liver, kidney, and lung has been reported in cats[47,48], consistent with our observation that *Flvcr1*, *Slc20a1*, and *Slc20a2* were highly expressed in stromal cells and distal convoluted tubule cells of the kidney, epicardial cells of the

heart, and ciliated cells of the lung. Notably, the expression proportions of *Slc20a1* and *Slc20a2* in corresponding cell types were much higher than that of *Flvcr1*. In addition to the commonly expressed cell types, we observed abundant expression of *Slc20a2* in heart cardiomyocytes and liver hepatocytes, whereas the expression of the other two receptors (*Slc20a1* and *Flvcr1*) was much less significant.

Furthermore, our study indicated that potential target tissue could be predicted based on viral receptor expression patterns. As a cell adhesion molecule that can mediate pseudorabies virus (PRV) entry, *Nectin1* colocalizes with E-cadherin at the adhesion junction of epithelial cells[49], consistent with the

enrichment of *Nectin1* in the epithelial cells of the eyelid. Interestingly, although PRV infection has been detected in the hearts of dogs, rats, and pigs[50–52], no PRV infection has been reported in cats heart previously. Here we observed that *Nectin1* was enriched in epicardial cells of feline heart, indicating cat heart could probably be permissive to PRV infection. In addition, *Anpep* is a receptor commonly used by feline infectious peritonitis virus, transmissible gastroenteritis virus, and infectious bronchitis virus. These viruses have been detected in kidney of cats[53,54]. Consistently, *Anpep* was found to be highly expressed in stromal cells and proximal tubule cells of kidney. Besides, we detected the enrichment of *Anpep* in hepatocytes and hepatic stellate cells of liver, indicating the possibility that these liver cells could be targeted by viruses using *Anpep* as entry factor (Fig. 4a).

SARS-CoV-2 entry factors are reportedly co-expressed with immunity genes in human and monkey cells[27,55], suggesting that SARS-CoV-2 target cells are conditioned to express immunity genes to reduce their susceptibility to viruses[27]. To explore whether this trend holds true for other viruses and non-model species, we applied Pearson's correlation coefficient analysis for all 13 feline viral receptors of all 31 cat cell types. Briefly, we identified the top 100 genes positively correlated with each virus receptor, followed by Gene Ontology (GO) enrichment analysis (Supplementary Data 4). We noticed that viral receptors co-expression analysis could reveal novel gene signatures that is not simply reflected in the cell type differentially expressed genes (DEGs) (Supplementary Fig. 6). Intriguingly, we observed co-expression of humoral immune response genes (*A2m*, *Fga*, and *C9*) with *Anpep* and *Nectin1*, co-expression of adaptive immune response genes (*Ptprc*, *Skap1*, and *Samsn1*) with *Cxcr4* and *Tfrc*, and co-expression of type 2 immune response genes (*Anxa1* and *Il33*) with *Nectin2* (Fig. 4c and Supplementary Data 4), suggesting the existence of extensive cell-virus cross-talk in the host–pathogen interaction interface. In summary, our data indicated that the co-expression of viral receptors and immunity genes may be more common than previously recognized. This phenomenon could be the result of the long-term "arms race" between viruses and hosts, which collectively shape the immune landscape of animals and host tropism of viruses[56].

Conservation of pulmonary cellular connectomes. In addition to exploring cell-virus cross-talk, a pan-species single-cell atlas can be applied to investigate divergent and conserved cell–cell interactions among multiple species. To identify putative cellular communications, we constructed a ligand-receptor-mediated interaction network for lung cells within each species, except the rabbit (due to limited cell numbers) and pigeon (due to limited number of ortholog genes), which revealed extensive and dynamic cellular cross-talk in pulmonary cells (Fig. 5a and Supplementary Data 5 and 6). Overall, the source connectome topology of distinct species was quite similar, with fibroblasts have dominant source weight and hub scores. As for target analysis of network centrality, endothelial cells function as signaling authority across most species including dog, hamster, lizard, goat and tiger. The immune cells were relatively less active, in both source and target analysis of connectome topology (Fig. 5b). This tendency was quite robust regardless of cutoff threshold values (Fig. 5c). In general, communication pairs were classified into 44 signaling modalities (Supplementary Data 6). Consistent with previous study[30], the AT1 cells dominated the vascular endothelial growth factor (VEGF) family signaling networks in most tested species (Fig. 5d).

We next identified pan-conserved cellular connectivity, which may correspond to ancient signaling vectors inherited from common ancestors of mammals, reptiles, and birds. In total, we detected 642 pairs of cell–cell interactions conserved among all nine species (Supplementary Data 6), most of which were associated with the extracellular matrix (collagens and fibronectin) and the development (VEGF) and morphogenesis (Semaphorin, SEMA) of pulmonary endothelial cells. Neuropilin-1, a receptor critical for vascular growth and remodeling, can convey SEMA and VEGF signaling[57]. Here we detected a fibroblast-Sema3c-Nrp1-endothelial and AT1-Vegfa-Nrp1-endothelial signaling axis in all nine species (Fig. 5e). Intriguingly, SEMA signaling plays a pivotal role in vascular patterning, and signal disruption can lead to anomalous pulmonary venous connections[58]. Here, cellular cross-talk mediated by other SEMA signaling components (e.g., Sema5a, 6a, and 6d) was also highly conserved (Supplementary Data 6).

In addition to pan-conserved cellular cross-talk, we explored why cell–cell cross-talk was conserved in all mammalian lung cells but was absent in the pulmonary cells of reptiles (lizard) and birds (duck). In total, we found 147 pairs of cellular interactions that were specifically conserved in mammals (Fig. 5f and Supplementary Data 6). The mammalian-specific conserved cellular connectivity included ADAM (Adam9-Itga3, Adam9-Itgav, and Adam9-Itgb5), ephrins (Efnb2-Pecam1 and Efna1-Epha7), fibronectins (Fn1-Itga3 and Fn1-Robo4), intracellular trafficking (Calr-Itga3 and Calr-Itgav), laminin signaling (Lama2, Lama4, Lamc1, and Lamc2 to Itga3), matrix metalloproteinases (MMPs) (Timp2-Itga3), NOTCH (Jag1-Notch2), SLIT (Slit2-Robo4), VEGF (Vegfc-Lyve1), and TGFB (Tgfb2-Eng). Jag1-Notch2 signaling is reported to play a major role in orchestrating alternative cell fate choices of ciliated cells[59]. Indeed, the Jag1 ligand broadcast from AT1 and AT2 cells is predicted to connect to the Notch2 receptor in ciliated cells. In addition, MMPs play critical roles in lung organogenesis[60]. In our study, Timp2 signaling from the fibroblasts and endothelial cells may target the Itga3 receptor of the AT1, AT2, ciliated, and endothelial cells. Endothelial ephrinB2 function was reported to be essential for lung alveolar formation[61]; here we detected endothelial cell-Efnb2-Pecam1-AT2 cell signaling was conserved in all mammalian species (pangolin, dog, hamster, deer, goat, cat, and tiger) but was absent in lizard and duck. Transforming growth factor-β plays crucial roles in epithelial–mesenchymal interactions for proper alveolarization and lung branching morphogenesis, and is thus critical for lung organogenesis and homeostasis[62]. Our data suggested that mammalian AT1 cells, AT2 cells, ciliated cells, and fibroblasts may regulate endothelial cells via Tgfb2-Eng signaling, but such signaling was not observed in the non-mammalian species. In summary, this study systematically revealed those highly conserved and lineage-specific cell-to-cell signaling within vertebrate lungs.

Conservation of regulomes in pulmonary cells. To reveal the regulatory mechanisms underlying alveolar development from the perspective of evolutionary biology, we predicted the regulomes in pulmonary cells for nine species (rabbit and pigeon were excluded for the same reasons mentioned above), resulting in an average of 469,796 Transcription Factor (TF)-target interactions for nine cell types (GENIE3 score > 0.01) (Supplementary Data 7–15). The number of TF-target interactions conserved in at least two species ranged from 5049 in AT2 cells to 44,009 in T cells (Fig. 6a and Supplementary Data 16). Encouragingly, several regulators for AT1 cells (Cux1 and Gata6)[63], AT2 cells (Etv5)[63], ciliated cells (Rfx3 and Glis3)[64,65], secretory cells (Nfib)[66],
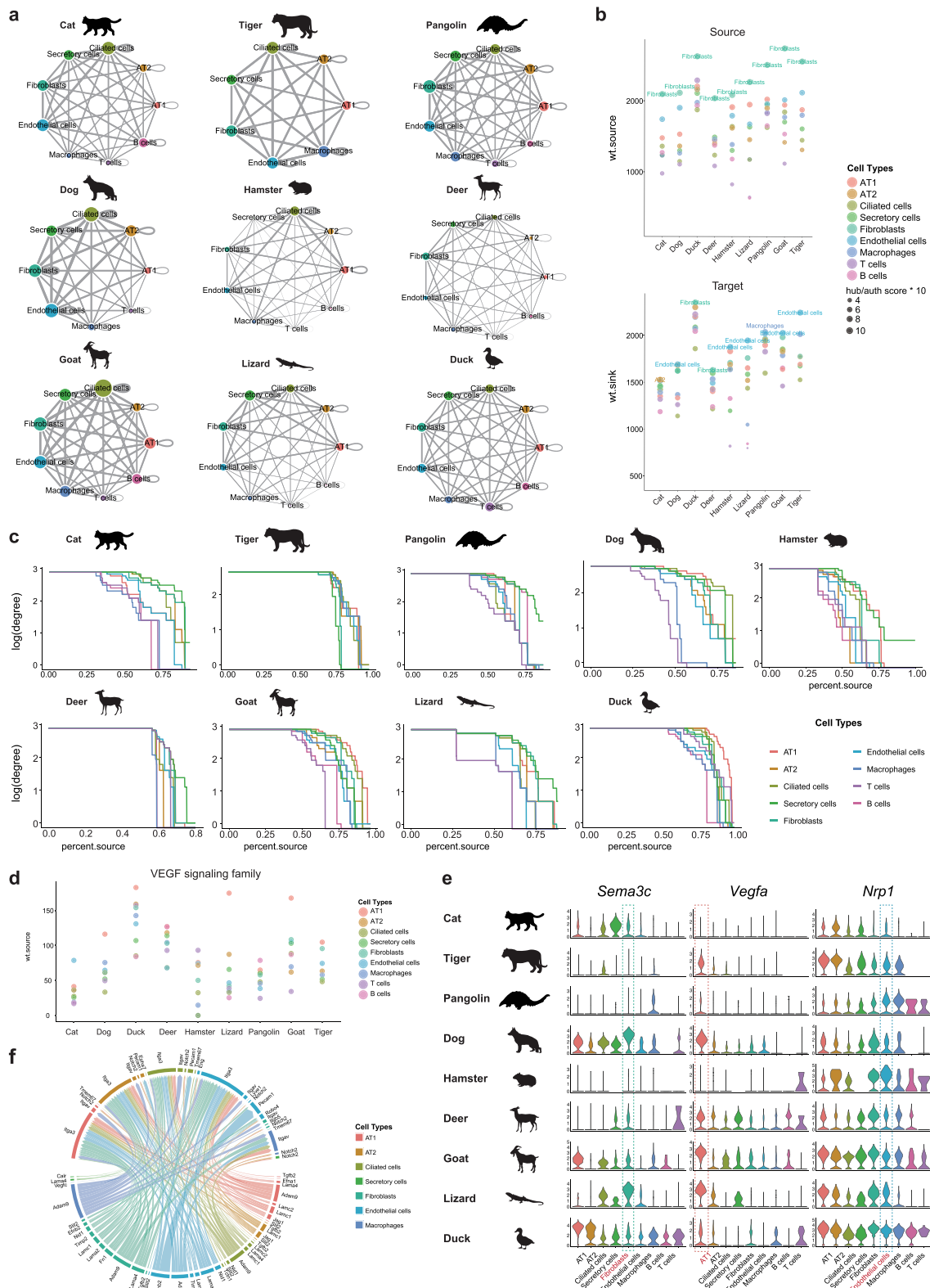
Fig. 5 Conserved pulmonary cellular connectomes. a Communication network of receptor–ligand pairs between lung cell types of indicated species. Cell type was represented by colored node, of which the size was directly proportional to sum of receptor–ligand pairs between this node and all other nodes. Edge width was directly proportional to number of receptor–ligand pairs between two connected nodes. **b** Plot (upper) showing source weight and hub score of various cell types in indicated nine species. Plot (lower) showing target weight and authority score of nine cell types. **c** Comparison of degree rankings and the effect of thresholding of lung cell types in indicated species. The *x*-axis ("percent.source") is percent of cell type expressing the marker, and the *y*-axis is the degree of each node plotted on a logarithmic scale. **d** Plot showing source weight of VEGF signaling family in nine cell types. **e** Violin plots showing expression level of indicated ligand and receptor genes within each cell type of lung in different species. **f** Circos plot of mammalian-specific conserved connectome in cat. Receptors and ligands were displayed near the upper and lower half circle, respectively.

fibroblasts (Zeb2 and Foxp1)[67,68], endothelial cells (Tcf12 and Dach1)[69,70], macrophages (Mef2a and PPARg)[71,72], T cells (Sox5 and Ikzf1)[73,74], and B cells (Mef2c and Ets1)[75,76] were active in the genetic regulatory network of the corresponding cell types (Supplementary Data 16), consistent with their expected regulatory functions. Of particular interest, we found a variety of regulatory circuits that were highly conserved among multiple species. Briefly, 37, 29, 533, 84, 151, 221, 1 097, 217 and 126 pairs of TF-target interactions were deeply conserved in AT1 cells, AT2 cells, ciliated cells, secretory cells, fibroblasts, endothelial cells, macrophages, T cells, and B cells, respectively (present in at least four species) (Supplementary Data 17). In the AT1 cells, 37 interactions associated with 12 TFs (Nfia, Dach1, Foxp1, Tox, Erg, Meis1 Cux1, Epas1, Mecom, Tcf12, Tead1, and Zeb2) were conserved in at least four of the nine species investigated. Notably, Cux1 contributed to 49% of the highly conserved interactions in AT1 cells (Fig. 6b). Cux1 is a known transcription factor regulating AT1 cell differentiation[63]. Rtkn2, the predicted target of Cux1, is reported to be an AT1 cell-specific gene[77]. The deeply conserved regulatory relationships between Cux1 and other functional genes (Lmo7, Cped1, Rtkn2, Ap1s3, Arhgef26, Cadm1, Col4a3, Ctdspl, Ctnna1, Lama3, Nrg1, Pdzd2, Unc13b, and Wwc2) suggest that Cux1 may manipulate the expression patterns of those developmental genes to instruct the differentiation, development, and morphogenesis of AT1 cells in a species-conserved manner.

In addition to well-known regulators, we also identified a variety of putative novel regulators in each cell type, e.g., Epas1, Klf6, Myb, Nfia, Tcf7l2, Tox, and Zeb2 in AT1 cells; Zbtb16, Sox5, and Tox in ciliated cells, and Mitf, Arid1b, Trps1, Etv6, Foxn3, Tcf12, Jarid2, Lrrfip1, Arid5b, Kmt2c, Elf1, and Mllt10 in macrophages. The regulatory functions of these transcription factors were inferred based on the enriched GO terms of their predicted target genes (Supplementary Data 18), which supported the proposed functions of each regulator in the corresponding cell type. For example, targets for AT1 regulators were closely related to GO terms covering epithelial cell development (Arhgef26, Cdk6, and Tmod1), lung epithelium development (Foxp2, Gata6, and Ncor2), morphogenesis of a branching epithelium (Greb1l, Cd44, and Dlg1), and epithelial tube morphogenesis (Tgfbr2, Dlc1, and Efnb2). Additionally, targets under the control of ciliated cell regulators were enriched in GO terms associated with cilium organization (Bbs9, Cfap43, and Rfx3), cilium assembly (Lrguk, Mak, and Rp1), cilium movement (Ccdc114, Ccdc39, and Dnah10), cilium-dependent cell motility (Cfap57, Spag16, and Tekt1), motile cilium assembly (Intu, Kif3a, and Lrrc46), epithelial cilium movement involved in determination of left, right asymmetry (Dnah11, Lrrc6, and Ccdc40), and regulation of cilium beat frequency (Armc4, Dnaaf1, and Bbs4). Targets for macrophage regulators were associated with macrophage chemotaxis (Mapk14, Lgals3, and Ptprj), macrophage migration (Rpl13a and Nup85), and macrophage cytokine production (Cd36, Cd74, and Sema7a) (Fig. 6c). Overall, our study systematically identified conserved regulators for pulmonary cells, including both well-recognized and novel regulators.

Conserved gene modules of AT1 and AT2 cells. Conserved core expression programs have been identified by cross-species comparison of microglia single-cell RNA sequencing data across evolutionary timescales[29]. Here we compared the transcriptomic landscapes of nine species (cat, tiger, pangolin, dog, hamster, deer, goat, lizard, and duck) spanning more than 312 million years of evolution[78]. Briefly, we selected 5442 genes showing expression in the lung cells of the nine species, followed by

hierarchical clustering (see "Methods" and Supplementary Data 19). Cluster 4, 5 (1432 genes) from AT1 cells (Fig. 7a and Supplementary Data 19) and cluster 3, 5 (1429 genes) from AT2 cells (Fig. 7b and Supplementary Data 19) showed conserved expression patterns across all nine species and were thus considered as conserved core gene expression programs for the AT1 and AT2 cells, respectively. Those cell type conserved genes showed largely non-overlapping molecular signature with cell type highly expressed genes (Supplementary Fig. 7). The AT1 cell core gene expression program was related to epithelial cell morphogenesis, proliferation, migration, and respiratory system development (Fig. 7c and Supplementary Data 20). In contrast, the AT2 cell-conserved genes were enriched in stem cell differentiation, stem cell population maintenance, activation of innate immune response, and lung alveolus development (Fig. 7d and Supplementary Data 20). For example, Cav1, a critical regulator of lung injury that is highly expressed in mouse AT1 cells[79,80], was conserved in the AT1 cells in multiple species (Fig. 7e), and Etv5, which is essential for the maintenance of mouse AT2 cells[81], was deeply conserved in the AT2 cells (Fig. 7f).

Interestingly, a set of lung disease genes overlapped in the core gene expression programs of the AT1 and AT2 cells, thus linking evolutionary biology with pulmonary pathology. For example, multiple genes in the AT1 cell program were closely related to pulmonary veno-occlusive disease (PVOD) (Bmpr2 and Eif2ak4), lung cancer (Itga9 and Braf), pulmonary fibrosis (Cadm1 and Itgb6), pulmonary hypertension (Bmpr2 and Cav1), and pulmonary emphysema (Itgb6). Several genes of AT2 core program were strongly associated with lung disease such as lung cancer (Mvp and Braf), pulmonary fibrosis (Parn), and PVOD (Eif2ak4) (Fig. 7a and Supplementary Data 20). Here we presented several examples that evolutionarily conserved genes are developmentally important, and interruptions in those genes can lead to disorders of well-balanced developmental programs, thus leading to a variety of pathological consequences. Collectively, we revealed core gene expression programs of alveolar cells that are highly conserved in mammals, reptiles, and birds, thus presenting gene modules of fundamental importance to alveolar cellular identity maintenance.

PANDORA: a website for non-model specie single-cell atlas. To fully exploit single-cell resources, we developed an online platform, named PANDORA (http://120.79.46.200:81/Pandora), composed of five main functional modules: i.e., panAtlas, panScan, panConnectome, panRegulome, and panCoreProgram. The panAtlas module allows users to explore gene expression patterns of any genes in any tissue included in this study. In the panScan module, visualization of the expression patterns of virus receptors in cell populations of each tissue can be calculated, giving clues about cell tropism of each virus. In the panConnectome module, cellular interactions mediated by ligands and receptors can be depicted. In the panRegulome module, putative TF–target regulations within each cell type can be predicted. In the panCoreProgram module, the core expression program for each cell type can be presented. In summary, we developed a comprehensive and integrated online platform, which can hopefully maximize the functionality and applicability of our resources.

Discussion
The SARS-CoV-2 pandemic has caused a catastrophic global health crisis. Although drugs and vaccines will help reduce loss in life and economic impact caused by COVID-19, their large-scale application is time-consuming and labor-intensive. Thus, the identification of intermediate hosts or animal reservoirs is crucial for the prevention and control of SARS-CoV-2 and other zoonotic viruses. So far, functional and genetic analyses of ACE2
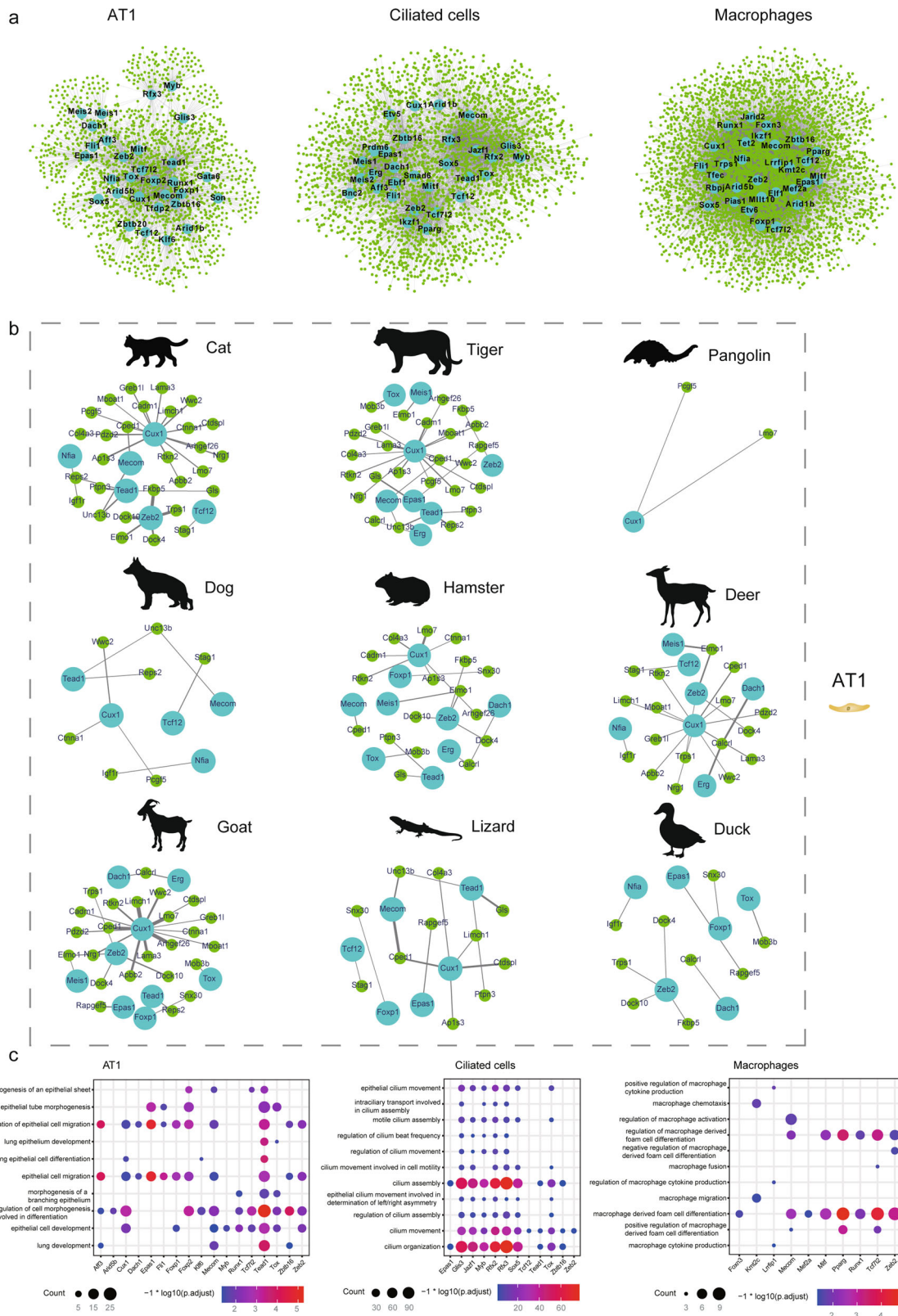
Fig. 6 Conservation of regulomes in pulmonary cells. a Genetic regulatory network constructed with top 30 TFs with largest number of interactions and predicted target genes in AT1 cells, ciliated cells, and macrophages. **b** Highly conserved genetic regulatory networks in AT1 cells within each indicated species. Light blue nodes represent regulators, green nodes represent corresponding target genes. Edge width is proportional to weight of regulation and node size is proportional to number of target genes of regulator. **c** GO enrichment related to cellular functions of predicted target genes of top 30 TFs in AT1 cells, ciliated cells, and macrophages, respectively. Dot color represents significance level of enrichment analysis and dot size is count of target genes classified in GO terms. P-values were calculated using hypergeometric test. Multiple comparisons adjustment was performed using Benjamini and Hochberg method. Exact P-value and source data were included in the Source Data file.
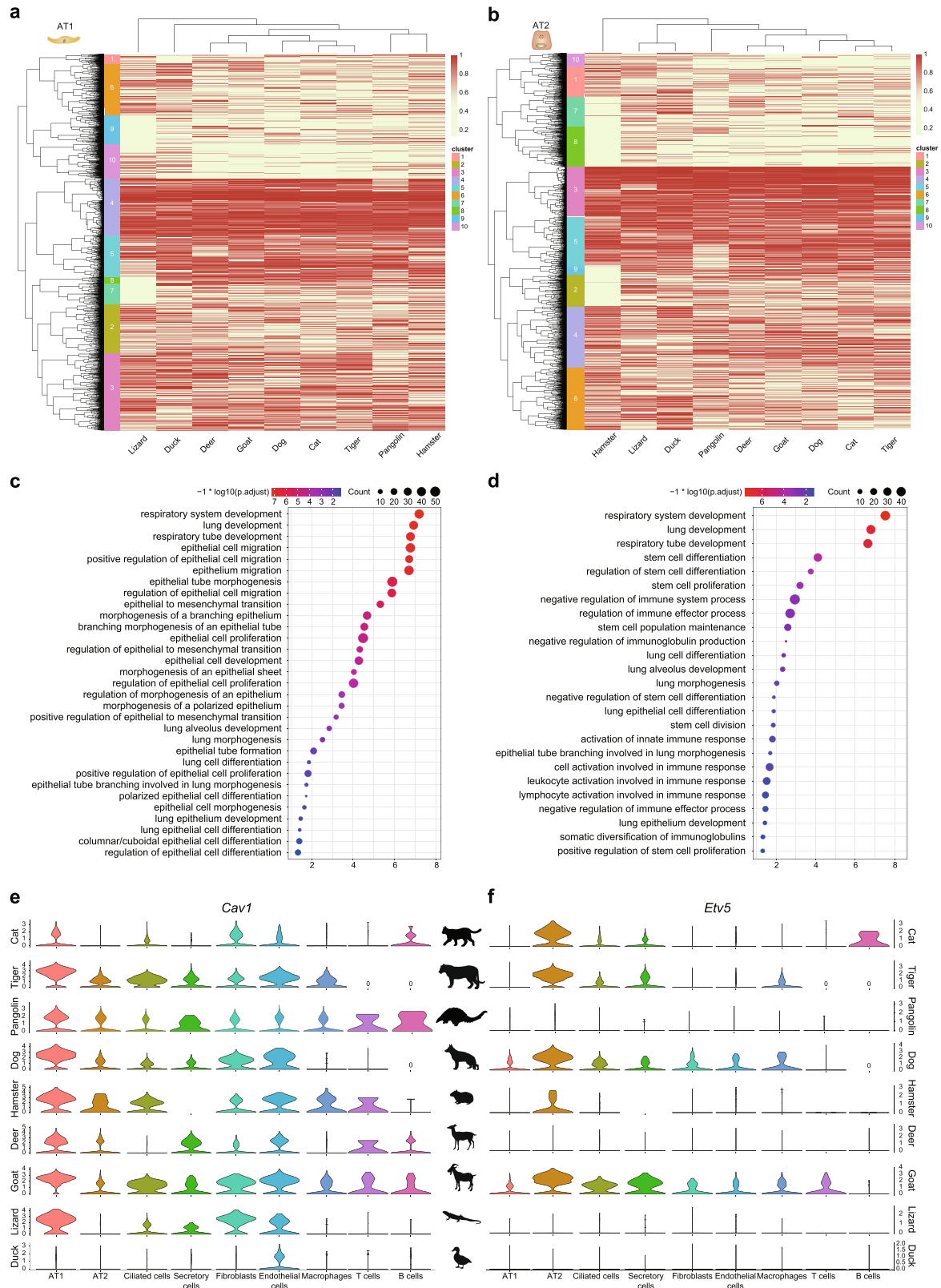
Fig. 7 Conserved core expression programs in AT1 and AT2 cells. a, b Heatmap showing clustering of orthologs in AT1 (**a**) and AT2 cells (**b**) across different species, with cluster index displayed on left. Each row represents one ortholog and each column represents one species. Colors represent expression levels of orthologs. **c, d** GO terms enriched in AT1 and AT2 cell-conserved orthologs. **e, f** Violin plot showing cell type-specific expression patterns of AT1 cell-conserved gene (*Cav1*) and AT2 cell-conserved gene (*Etv5*). *P*-values were calculated using hypergeometric test. Multiple comparisons adjustment was performed using Benjamini and Hochberg method. Exact *P*-value and source data were included in the Source Data file.

orthologs in 48 species have found that 44 can mediate cell entry of SARS-CoV-2, implicating a broad range of host tropism for this virus[27]. With the advancement of single-cell sequencing technologies, the single-cell atlas for human, mouse, zebrafish, fruit fly, frog, and nematode have been successfully constructed[20–23,82,83]. However, studies on the cell composition of non-model species are lacking. To expand single-cell resources for non-model species, we generated a single-cell atlas for 11 species that are in close contact with humans (livestock, poultry, and pets) or possibly harbor enzootic viruses, thus producing a transcriptome of 272,148 high-quality cells derived from 29 tissues. Based on these resources, we screened putative virus target cells, revealed critical regulators of cell fate commitment, dissected genetic regulatory networks of pulmonary cells, and identified core gene expression programs deeply conserved among mammals, reptiles, and birds.

SARS-CoV-2 target cells were widely distributed among tissues within the digestive system (esophagus, rectum), respiratory system (lung), and urinary system (kidney) of the cat. Regarding the licking habits of felines, cats could be infected by SARS-CoV-2 via the fecal-oral route or the airborne transmission route. Here, our results provide some explanation for the observation that cats are highly permissive to SARS-CoV-2[40] and highlight the necessity to monitor and evaluate the possible roles of cats during the COVID-19 pandemic.

Cellular signaling is crucial for the emergence of tissue properties and the maintenance of tissue homeostasis[30]. In this study, we painted a system-level portrait of pulmonary cell–cell cross-talk. In addition, we identified a variety of conserved cellular connectivity exclusively present in mammals. The evolution of vertebrate lungs suggests that the emergence of new pulmonary structures is usually derived from modifications of preceding ones in order to adapt to new functional requirements[84]. Yet, there remains a huge gap between lung morphological innovations and pulmonary cellular communication modifications. Thus, our study could provide some clues regarding the underlying mechanisms of evolution innovation from primitive lungs to advanced mammalian lungs at the cellular and molecular levels.

Several well-known master regulators of pulmonary cells were found to be active in corresponding cell types in a species-conserved manner, in line with their expected functions, thus providing a repertoire of regulatory circuits that deserve further investigation. In addition, we identified a range of novel regulators that may work cooperatively with canonical TFs to regulate pulmonary cell development. However, the functions of TFs in lung development of non-model species remain poorly understood, and follow-up functional studies are needed to reveal their previously underappreciated regulatory functions. Our study also showed that developmentally important regulators were evolutionarily conserved, indicating that conserved TF–target pairs may represent critical regulatory circuits, which are worthy of experimental verification. Thus, cross-species comparison of genetic regulatory networks among evolutionarily distant species offers a strategy to pinpoint regulatory circuits that are fundamental for lineage commitment and cell identity maintenance.

Taken together, our study provides a unique resource and framework to study putative virus target cells, making it possible to screen the entry factors of a wide range of viruses in various species in an unbiased manner. Evolutionarily conserved connectomes and regulomes are of fundamental importance for organism development and cell fate commitment. Considering the high diversity and heterogenicity of animal tissues, it is necessary to elucidate tissue evolution at single-cell resolution.

With the development of single-cell sequencing and advancement of international collaborative projects, atlases for more species will be generated at an accelerated speed. We anticipate that the information gained from the present study could augment future research and provide insights into prevention and control strategies against enzootic viruses such as SARS-CoV-2. Furthermore, our research should allow further investigations on the regulatory networks, cellular cross-talk, and core gene expression programs across evolutionary timescales. Hopefully, our resources can be utilized by evolutionary biology researchers to elucidate the cellular and molecular mechanisms underlying the origin and evolution of lungs, an important interface for pathogen–host interactions and a crucial respiratory organ that plays an important role in vertebrate adaptation to unique ecological niches.

There are several limitations in this study. First, the *ACE2/TMPRSS2*-expressing cells might be under-detected due to intrinsic feature of single-cell sequencing. Second, there may exist inter-individual heterogeneity of each species. Third, differences in innate and adaptive immune responses could putatively affect viral replication, assembly, and/or release[85]. Therefore, the results of this study should be interpreted cautiously and the exact host range for SARS-CoV-2 should be determined by experimental confirmation.

Methods

Ethics statement. All experimental procedures and sample collection protocols were performed with the approval of the Institutional Review Board on Ethics Committee of BGI (NO. BGI-IRB A20008). All sampling procedures strictly followed the "Guidelines on the Ethical Treatment of Experimental Animals" established by the Ministry of Science and Technology, China.

Sample collection. This study was fully endorsed and approved by Institutional Review Board on Ethics Committee of BGI. All sampling activities were conducted in collaboration with technical staff. All the samples were stored carefully and used for research purposes only. We declare that this study complied with the Convention on Biological Diversity and the Convention on Trade in Endangered Species of Wild Fauna and Flora. Tissue samples were obtained from 11 animals, including: (1) four pet species: cat (*Felis catus*, domestic short-haired cat), dog (*Canis lupus familiaris*, German Shepherd), hamster (*Mesocricetus auratus*, golden hamster), and lizard (*Anolis carolinensis*, green anoles); (2) two livestock species: goat (*Capra aegagrus hircus*, domestic goat) and rabbit (*Oryctolagus cuniculus domesticus*, domestic rabbit); (3) two poultry species: duck (*Anas platyrhynchos domesticus*, domestic duck) and pigeon (*Columba livia domestica*, domestic pigeon); and (4) three wild animal species: tiger (*Panthera tigris altaica*, Siberian tiger), pangolin (*Manis javanica*, Sunda pangolin), and deer (*Cervus nippon*, Sika deer). The pangolin samples were collected from an individual that died of natural causes in Guangdong Provincial Wildlife Rescue Center, China, and was immediately stored in a −80 °C freezer after dissection. The tiger samples were obtained from an individual that died of natural causes in the Siberian Tiger Park in Heilongjiang Province, China. Tiger samples were immediately stored in a −80 °C freezer after dissection. Deer samples and corresponding genome assembly were kindly provided by the Institute of Special Animal and Plant Sciences (ISAPS) of the Chinese Academy of Agricultural Sciences. Other animals were obtained from markets with permission from the BGI Ethics Committee. Animals were kept in a pathogen-free environment and provided with sufficient living space and adequate food and water. After animal euthanasia in accordance with the animal experiment guidelines issued by the Chinese Ministry of Science and Technology, dissection was carried out to separate each tissue. The collected tissues were rinsed using 1× phosphate-buffered saline, then quick-frozen, and stored in liquid nitrogen. Nuclei of each tissue were separated by mechanical extraction. Briefly, the tissues were first thawed, infiltrated by 1× homogenization buffer (containing 30 mM CaCl₂, 18 mM Mg(Ac)₂, 60 mM Tris-HCl pH 7.8, 320 mM sucrose, 0.1% NP-40, 0.1 mM EDTA, and 0.2 U/μl RNase inhibitor), and cut into smaller pieces, with the single nucleus then isolated by 2 ml of Dounce homogenizer. After filtration with a 30 μm strainer, the nuclear extraction was resuspended in 1% bovine serum albumin (BSA) containing 0.2 U/μl RNase inhibitor and centrifuged at $500 \times g$ for 10 min at 4 °C to discard cellular impurities within the supernatant. This step was repeated twice, after which the nuclei were recollected with 0.1% BSA containing 0.2 U/μl RNase inhibitor. Subsequently, 4′,6-diamidino-2-phenylindole was used to stain the nuclei and nucleus density was calculated under a fluorescence microscope for subsequent library construction.

Single-nucleus library construction and sequencing. The separated single nuclei of different tissues (e.g., lungs of cat, tiger, pangolin, dog, hamster, deer, rabbit, goat, lizard, duck, and pigeon; cat kidney, liver, heart, eyelid, esophagus, and rectum; tiger kidney, liver, spleen, and heart; pangolin kidney, liver, spleen, heart, esophagus, stomach, duodenum, and large intestine) underwent library construction using a Chromium Single-cell 3' GEM, Library & Gel Bead Kit v3 (PN-1000075) following the guidelines provided by the manufacturer. After conversion using the MGI Easy Universal DNA Library Preparation Reagent Kit, the libraries were sequenced using a compatible BGISEQ-500 sequencing platform.

Cross-species homolog gene conversion. Gene models were downloaded from the National Center for Biotechnology Information (NCBI) database (Supplementary Data 1). To facilitate integration of the cross-species single-cell lung data sets, we converted genes from other species to the mouse homologs. First, we downloaded the homologs of eight species (cat, tiger, dog, hamster, goat, rabbit, lizard, and duck) and the mouse using BioMart[86]. For the pangolin, pigeon, and deer, which lack homolog records on Ensemble. Single-copy orthologs were identified from two species genomes by cluster analysis of gene families using OrthoFinder[87] v2.3.3 with default parameters. Single-copy genes were extracted from the OrthoFinder output file. If a 1 : 1 match existed between a non-mouse and mouse gene, the non-mouse gene name was converted to the mouse gene name.

Single-cell transcriptomic data processing. Sequencing data were filtered using a custom script and the gene expression matrix was obtained using Cell Ranger v3.0.2 (10× Genomics). The genomes used for read alignment were downloaded from the NCBI Assembly (Supplementary Data 1). Single-cell analysis was conducted using Seurat[33,34]. Briefly, quality control was performed based on the following criteria: cells with mapped number of genes <200 or with mitochondrial percentage higher than 10% were removed. Variable genes were determined using Seurat's FindVariableGenes function with default parameters (selection.method = "vst", nfeatures = 2000). Clusters were identified via the FindClusters function (resolution = 1) in Seurat using principal components with a P-value < 0.01 and subsequently visualized using the RunTSNE function (reduction = "pca"). All DEGs for each Seurat object were identified using the FindAllMarkers function (only.pos = T, min.pct = 0.1, logfc.threshold = 0.25).

Collection of cell type markers. Cell type markers for each cell type were collected from previously published literature[30,88–91] and the CellMarker database[92].

Cell type identity inference. In general, lung cells were annotated based on the overall transcriptomic similarity with the reference data set and were combined with canonic cell type markers. Each lung atlas data set from the 11 species was integrated with the reference lung atlas (generated using 10× Genomics for human, pig, mouse, and rat)[30] to infer their putative identity. Briefly, data sets of lungs from different species were integrated with the reference data set using the Seurat FindIntegrationAnchors and IntegrateData functions with features after homolog conversion[33,34]. Subsequently, cells in each cluster were marked according to their origin (from custom or reference data). The cell type identity tag of the reference cells was obtained from the metadata of the reference lung data set and the proportion of each reference cell type (AT1 cells, AT2 cells, ciliated cells, secretory cells, fibroblasts, endothelial cells, macrophages, T cells, and B cells) in every cluster was calculated. The cell type contributing to the highest proportion of the corresponding cluster was adopted as the potential cell type of that cluster. Cell type identity was further confirmed according to the expression of canonical cell type markers. Regarding other tissues (heart, liver, spleen, and kidney), cells were annotated according to canonical markers.

GO term enrichment analysis. The gene list converted to mouse homologs was subjected to GO analysis using the hypergeometric test implemented in the clusterProfiler[93] package. GO enrichment level was evaluated by adjusted P-values and multiple test adjustment was conducted using the Benjamini–Hochberg (BH) method.

Screening of SARS-CoV-2 target cells. A cell was considered a SARS-CoV-2 target cell if *ACE2* and *TMPRSS2* expression levels were detected (unique molecular identifier, UMI > 0) simultaneously. The percentage of *ACE2&TMPRSS2*-positive cells was calculated for each cell type.

Pearson correlation coefficients. Pearson correlation analysis was conducted using the corr.test function in the psych package with the parameter: method = "pearson". Multiple test adjustment was conducted using the p.adjust function with parameter: method = "BH".

Screening of viral receptors in cat cells. Viruses that may infect cats were collected from previously published literature[41,45,46,53] and reviewed by virology experts. A receptor list of viruses was downloaded from the viralReceptor database (http://www.computationalbiology.cn:5000/viralReceptor). First, the percentage of

receptor-positive cells was calculated for each cell type, resulting in a receptor percentage matrix (receptors in columns, cell type in rows). We next performed principal component analysis on all cell types using the receptor percentage matrix. Pearson's correlation coefficients between receptors and all protein-coding genes were calculated, and the top 100 correlated genes of each receptor were subjected to GO term enrichment analysis.

Statistics analysis for *ACE2* and *TMPRSS2* co-expression enrichment. To test the significance of SARS-CoV-2 putative target cells in cat compared to other species, we randomly sampled 100 times of ciliated cells in each species, with 50% cells selected each time. After calculation of *ACE2* and *TMPRSS2* co-expression ratio in each sampling, Wilcoxon's rank sum test was conducted using the 100 co-expression ratios between cat and that of the other 10 species respectively. Similarly, we tested the significance of *ACE2* and *TMPRSS2* co-expression in ciliated cells, compared to macrophages.

Cellular communication analysis. Ligands and receptors were downloaded from the FANTOM5 database[94], a widely used database for inferring connectome in various species[30,95–97], and included all literature-supported mouse ligands and receptors. Connectome networks were constructed according to the expression of ligands and receptors and both degree and centrality were calculated based on interactions passing different percentage thresholds using methods described by Raredon[30]. Nodal degree, hub authority, and centrality were calculated using the R package igraph[98]. The threshold of cognate ligand-receptor pair expression was greater than 5% of cells in source and target cell types. We next applied network centrality analysis and mode dominate analysis to connectomes using the CompareCentrality functions in the R package Connectome (https://github.com/msraredon/Connectome). CircosPlot function was used to visualize mammalian-specific conserved interactions.

Identification of pan-conserved and mammalian-specific conserved cellular interactions. A signaling axis was defined as a combination of the following four components: source cells, ligands, receptors, target cells. If a signaling axis was present in all nine species investigated (cat, tiger, pangolin, dog, hamster, deer, goat, lizard, and duck), then it was considered as a pan-conserved cellular interaction. In addition, if a signaling axis was present in all seven mammalian species (cat, tiger, pangolin, dog, hamster, deer, and goat) but absent in the lizard and duck, it was defined as a mammalian-specific conserved cellular interaction.

TF–target interaction inference. The TF list for mouse (*Mus musculus*) was downloaded from the animalTFDB3.0[99]. TFs and other genes were filtered out if expressed in <5% of corresponding cell types. Only protein-coding genes were maintained for subsequent analysis. Briefly, GENIE3[100] was employed to predict putative regulatory circuits, with TF–target interactions with a weight value ≥0.01 retained for further regulatory network construction. Regulomes of representative TFs were visualized using Cytoscape and the R package igraph[98,101].

Conservation analysis of TF–target pairs. The total frequency of each TF–target pair in a specific cell type in the nine species was counted to evaluate its conservation level. If a given TF–target pair was present in at least two of the same cell types in the nine species, it was regarded as "conserved." If the count was ≥4, then the TF–target pair was considered as "highly conserved."

Identification of putative cell type regulators. Cell type regulators were identified as follows: (1) TF–target interactions needed to be present in a specific cell type in at least two species. (2) TFs in conserved TF interactions (in at least two species) in each cell type were ranked in descending order according to the number of targets under their regulation. (3) Targets for each of the top 30 TFs in each cell type were subjected to GO term enrichment. (4) TFs with targets enriched in GO terms closely associated with expected biological processes of corresponding cell types were considered as putative regulators. (5) Extensive literature mining was performed for the putative regulators identified in the last step. If any report supported a clear link between the investigated TF and corresponding cell type, then it was considered as a known regulator. Otherwise, it was annotated as a putative novel regulator.

Conserved core expression programs identification. A gene was considered as expressed in a species if it was expressed in at least one cell type (UMI ≥ 1) in the lung atlas of that species. We intersected the expressed gene list of the nine species and obtained a set of 5442 commonly expressed genes for subsequent analysis. We first calculated the average gene expression of each cell type among the nine species. To obtain the percentile values of genes in a certain cell type of a particular species, we sorted genes according to their average expression levels from low to high, and then divided the gene ranking by the total number of genes. Using the same strategy, we generated a percentile matrix for each cell type of all nine species. Hierarchical clustering was applied to the percentile matrices in the AT1 and AT2 cells in all nine species. The clustered genes were ranked by average percentile

values, and the top two clusters were considered the core gene expression programs for the AT1 and AT2 cells.

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

Data availability

The raw data and processed data generated in this study have been deposited in the NCBI database under accession code PRJNA747757 and GEO database under accession code GSE183300. Alternatively, raw transcriptome sequencing data and processed data were deposited at the CNSA (CNGB Nucleotide Sequence Archive) under accession number CNP0001882 and CNP0001889. The single-cell atlases of all investigated species in this study are available via http://120.79.46.200:81/Pandora/Download.html. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

Custom scripts for data analysis in this study were present in https://github.com/DongshengChen-TY/Atlas (https://doi.org/10.5281/zenodo.5139392).

References

1. Gorbalenya, A. et al. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. *Nat. Microbiol.* https://doi.org/10.1101/2020.02.07.937862 (2020).
2. World Health Organization. *WHO/Europe | Coronavirus disease (COVID-19) Outbreak - WHO Announces COVID-19 Outbreak a Pandemic* (WHO/Europe, 2020).
3. Director-General, W. *WHO Director-General's Remarks at the Media Briefing on 2019-nCoV on 11 February 2020* (World Health Organization, 2020).
4. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature*, https://doi.org/10.1038/s41586-020-2008-3 (2020).
5. Guan, W. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2002032 (2020).
6. Corman, V. M. et al. Evidence for an ancestral association of human coronavirus 229E with bats. *J. Virol.* https://doi.org/10.1128/jvi.01755-15 (2015).
7. Corman, V. M., Muth, D., Niemeyer, D. & Drosten, C. In *Advances in Virus Research*. 100, 163–188 (Academic, 2018).
8. Huynh, J. et al. Evidence supporting a zoonotic origin of human coronavirus strain NL63. *J. Virol.* https://doi.org/10.1128/jvi.00906-12 (2012).
9. Lau, S. K. P. et al. Discovery of a novel coronavirus, China rattus coronavirus HKU24, from Norway rats supports the murine origin of Betacoronavirus 1 and has implications for the ancestor of betacoronavirus lineage A. *J. Virol.* https://doi.org/10.1128/jvi.02420-14 (2015).
10. Moreno, A. et al. Detection and full genome characterization of two beta CoV viruses related to Middle East respiratory syndrome from bats in Italy. *Virol. J.* https://doi.org/10.1186/s12985-017-0907-1 (2017).
11. Yang, X.-L. et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *J. Virol.* https://doi.org/10.1128/jvi.02582-15 (2016).
12. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
13. Lam, T. T. Y. et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, https://doi.org/10.1038/s41586-020-2169-0 (2020).
14. Liu, P., Chen, W. & Chen, J. P. Viral metagenomics revealed sendai virus and coronavirus infection of malayan pangolins (*Manis javanica*). *Viruses*, https://doi.org/10.3390/v11110979 (2019).
15. Wong, M., Javornik Cregeen, S., Ajami, N. & Petrosino, J. Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019. Preprint at *bioRxiv*, https://doi.org/10.1101/2020.02.07.939207 (2020).
16. Sit, T. H. C. et al. Infection of dogs with SARS-CoV-2. *Nature* 1–6 https://doi.org/10.1038/s41586-020-2334-5 (2020).
17. Zhang, Q. et al. SARS-CoV-2 neutralizing serum antibodies in cats: a serological investigation. Preprint at *bioRxiv*, https://doi.org/10.1101/2020.04.01.021196 (2020).
18. Hofmann, H. et al. Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. USA*, https://doi.org/10.1073/pnas.0409465102 (2005).
19. Li, W. et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, https://doi.org/10.1038/nature02145 (2003).
20. Han, X. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).
21. Han, X. et al. Mapping the mouse cell atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).
22. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
23. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).
24. Zhang, L. et al. A high-resolution cell atlas of the domestic pig lung and an online platform for exploring lung single-cell data. *J. Genet. Genomics* **48**, 411–425 (2021).
25. Zhu, J. et al. Single-cell atlas of domestic pig cerebral cortex and hypothalamus. *Sci. Bull.* **66**, 1448–1461 (2021).
26. Zhong, S. et al. Single-cell brain atlas of Parkinson's disease mouse model. *J. Genet. Genomics* **48**, 277–288 (2021).
27. Sungnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
28. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
29. Geirsdottir, L. et al. Cross-species single-cell analysis reveals divergence of the primate microglia program. *Cell* **179**, 1609–1622.e16 (2019).
30. Raredon, M. S. B. et al. Single-cell connectomic analysis of adult mammalian lungs. *Sci. Adv.* **5**, eaaw3851 (2019).
31. Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* 1–7 https://doi.org/10.1038/s41586-020-2313-x (2020).
32. McAloose, D. et al. From people to *Panthera*: natural SARS-CoV-2 infection in tigers and lions at the Bronx Zoo. Preprint at *bioRxiv*, https://doi.org/10.1101/2020.07.22.213959 (2020).
33. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
34. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell*, https://doi.org/10.1016/j.cell.2019.05.031 (2019).
35. Ziegler, C. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is enriched in specific cell subsets across tissues. *SSRN Electronic J.* https://doi.org/10.2139/ssrn.3555145 (2020).
36. Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* 1–12 https://doi.org/10.1038/s41577-020-0311-8 (2020).
37. Damas, J. et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl Acad. Sci. USA* **117**, 202010146 (2020).
38. Liu, Y. et al. Functional and genetic analysis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2. Preprint at *bioRxiv*, https://doi.org/10.1101/2020.04.22.046565 (2020).
39. Sia, S. F. et al. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* 1–7 https://doi.org/10.1038/s41586-020-2342-5 (2020).
40. Shi, J. et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. *Science* eabb7015 https://doi.org/10.1126/science.abb7015 (2020).
41. da Silva, R. A., de Souza, A. M. & André, C. A. [Studies on rabies in Brazil. II. Isolation of the rabies virus from the lungs, kidney, bladder and tissues of other organs of the naturally infected domestic cat]. *Bol. Oficina Sanit. Panam.* **64**, 506–507 (1968).
42. Alexander, R. et al. [Pathogenesis of experimental rabies infection in foxes and cats infected with viruses isolated from hamsters (*Cricetus cricetus*)]. *Vet. Med.* **26**, 233–246 (1981).
43. Guo, Y. et al. Early events in rabies virus infection—attachment, entry, and intracellular trafficking. *Virus Res.* **263**, 217–225 (2019).
44. Frey, S. C. S., Hoover, E. A. & Mullins, J. I. Feline immunodeficiency virus cell entry. *J. Virol.* **75**, 5433–5440 (2001).
45. Sugai, J. et al. Identification of envelope determinants of feline leukemia virus subgroup B that permit infection and gene transfer to cells expressing human Pit1 or Pit2. *J. Virol.* **75**, 6841–6849 (2001).
46. Brown, J. K., Fung, C. & Tailor, C. S. Comprehensive mapping of receptor-functioning domains in feline leukemia virus subgroup C receptor FLVCR1. *J. Virol.* **80**, 1742–1751 (2006).
47. Helfer-Hungerbuehler, A. K. et al. Long-term follow up of feline leukemia virus infection and characterization of viral RNA loads using molecular methods in tissues of cats with different infection outcomes. *Virus Res.* **197**, 137–150 (2015).
48. Leite-Filho, R. V., Panziera, W., Bandinelli, M. B. & Pavarini, S. P. Pathological characterization of lymphoma with pulmonary involvement in cats. *J. Comp. Pathol.* **165**, 6–12 (2018).
49. Yoon, M. & Spear, P. G. Disruption of adherens junctions liberates Nectin-1 to serve as receptor for herpes simplex virus and pseudorabies virus entry. *J. Virol.* **76**, 7203–7208 (2002).

50. Olson, G. R. & Miller, L. D. Studies on the pathogenesis of heart lesions in dogs infected with pseudorabies virus. *Can. J. Vet. Res.* **50**, 245–250 (1986).

51. Ter Horst, G. J. Central autonomic control of the heart, angina, and pathogenic mechanisms of post-myocardial infarction depression. *Eur. J. Morphol.* **37**, 257–266 (1999).

52. Irnaten, M. et al. Activity of cardiorespiratory networks revealed by transsynaptic virus expressing GFP. *J. Neurophysiol.* **85**, 435–438 (2001).

53. Evermann, J. F., Baumgartener, L., Ott, R. L., Davis, E. V. & Mckeirnan, A. J. Characterization of a feline infectious peritonitis virus isolate. *Vet. Pathol.* **18**, 256–265 (1981).

54. Mochizuki, M. & Furukawa, H. An enzyme-linked immunosorbent assay using canine coronavirus-infected CRFK cells as antigen for detection of anti-coronavirus antibody in cat. *Comp. Immunol. Microbiol. Infect. Dis.* **12**, 139–146 (1989).

55. Han, L. et al. Single-cell atlas of a non-human primate reveals new pathogenic mechanisms of COVID-19. Preprint at bioRxiv, https://doi.org/10.1101/2020.04.10.022103 (2020).

56. Simmonds, P., Aiewsakun, P. & Katzourakis, A. Prisoners of war — host adaptation and its constraints on virus evolution. *Nat. Rev. Microbiol.* **17**, 321–328 (2019).

57. Gu, C. et al. Neuropilin-1 conveys semaphorin and VEGF signaling during neural and cardiovascular development. *Dev. Cell* **5**, 45–57 (2003).

58. Degenhardt, K. et al. Semaphorin 3d signaling defects are associated with anomalous pulmonary venous connections. *Nat. Med.* **19**, 760–765 (2013).

59. Kiyokawa, H. & Morimoto, M. Notch signaling in the mammalian respiratory system, specifically the trachea and lungs, in development, homeostasis, regeneration, and disease. *Dev. Growth Differ.* **62**, 67–79 (2020).

60. Greenlee, K. J., Werb, Z. & Kheradmand, F. Matrix metalloproteinases in lung: Multiple, multifarious, and multifaceted. *Physiol. Rev.* **87**, 69–98 (2007).

61. Wilkinson, G. A., Schittny, J. C., Reinhardt, D. P. & Klein, R. Role for ephrinB2 in postnatal lung alveolar development and elastic matrix integrity. *Dev. Dyn.* **237**, 2220–2234 (2008).

62. Saito, A., Horie, M. & Nagase, T. TGF-β signaling in lung health and disease. *Int. J. Mol. Sci.* **19**, 2460 (2018).

63. Whitsett, J. A., Kalin, T. V., Xu, Y. & Kalinichenko, V. V. Building and regenerating the lung cell by cell. *Physiol. Rev.* **99**, 513–554 (2019).

64. Kang, H. S., Beak, J. Y., Kim, Y.-S., Herbert, R. & Jetten, A. M. Glis3 is associated with primary cilia and Wwtr1/TAZ and implicated in polycystic kidney disease. *Mol. Cell. Biol.* **29**, 2556–2569 (2009).

65. Didon, L. et al. RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium. *Respir. Res.* **14**, 70 (2013).

66. Hsu, Y. C. et al. Mesenchymal nuclear factor I B regulates cell proliferation and epithelial differentiation during lung maturation. *Dev. Biol.* **354**, 242–252 (2011).

67. Shao, X. & Wei, X. FOXP1 enhances fibrosis via activating Wnt/β-catenin signaling pathway in endometriosis. *Am. J. Transl. Res.* **10**, 3610–3618 (2018).

68. Meikle, Z., Yu, L., Netticadan, T. & Wigle, J. The regulation of fibroblast to myofibroblast activation by zinc finger E-box binding homeobox 1 and 2. *FASEB J.* **34**, 1–1 (2020).

69. Chang, A. H. et al. DACH1 stimulates shear stress-guided endothelial cell migration and coronary artery growth through the CXCL12-CXCR4 signaling axis. *Genes Dev.* **31**, 1308–1324 (2017).

70. Li, Y. et al. Targeted disruption of TCF12 reveals HEB as essential in human mesodermal specification and hematopoiesis. *Stem Cell Rep.* **9**, 779–795 (2017).

71. Aude-Garcia, C. et al. Dual roles for MEF2A and MEF2D during human macrophage terminal differentiation and c-Jun expression. *Biochem. J.* **430**, 237–244 (2010).

72. Bouhlel, M. A. et al. PPARγ activation primes human monocytes into alternative M2 macrophages with anti-inflammatory properties. *Cell Metab.* **6**, 137–143 (2007).

73. Powell, M. D., Read, K. A., Sreekumar, B. K. & Oestreich, K. J. Ikaros zinc finger transcription factors: Regulators of cytokine signaling pathways and CD4+ T helper cell differentiation. *Front. Immunol.* **10**, 1299 (2019).

74. Suto, A., Tanaka, S. & Nakajima, H. Sox5 and Th17 cell differentiation. *Oncotarget* **6**, 19952–19953 (2015).

75. Sunshine, A. et al. Ets1 controls the development of B cell autoimmune responses in a cell-intrinsic manner. *ImmunoHorizons* **3**, 331–340 (2019).

76. Wilker, P. R. et al. Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation. *Nat. Immunol.* **9**, 603–612 (2008).

77. Little, D. R. et al. Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. *Proc. Natl Acad. Sci. USA* **116**, 20545–20555 (2019).

78. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

79. Maniatis, N. A., Chernaya, O., Shinin, V. & Minshall, R. D. Caveolins and lung function. *Adv. Exp. Med. Biol.* **729**, 157–179 (2012).

80. Jin, Y., Lee, S. J., Minshall, R. D. & Choi, A. M. K. Caveolin-1: a critical regulator of lung injury. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **300**, L151 (2011).

81. Zhang, Z. et al. Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. *Proc. Natl Acad. Sci. USA* **114**, 3903–3908 (2017).

82. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).

83. Davie, K. et al. A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell* **174**, 982–998.e20 (2018).

84. Liem, K. F. Form and function of lungs: the evolution of air breathing mechanisms. *Integr. Comp. Biol.* **28**, 739–759 (1988).

85. Douam, F. et al. Genetic dissection of the host tropism of human-tropic pathogens. *Annu. Rev. Genet.* **49**, 21–45 (2015).

86. Cunningham, F. et al. Ensembl 2019. *Nucleic Acids Res* **47**, D745–D751 (2019).

87. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

88. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single cell RNA sequencing. Preprint at bioRxiv, https://doi.org/10.1101/742320 (2020).

89. Wang, P. et al. Dissecting the global dynamic molecular profiles of human fetal kidney development by single-cell RNA sequencing. *Cell Rep.* https://doi.org/10.1016/j.celrep.2018.08.056 (2018).

90. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* https://doi.org/10.1038/s41467-018-06318-7 (2018).

91. Crinier, A. et al. High-dimensional single-cell analysis identifies organ-specific signatures and conserved NK cell subsets in humans and mice. *Immunity*, https://doi.org/10.1016/j.immuni.2018.09.009 (2018).

92. Zhang, X. et al. CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gky900 (2019).

93. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

94. Ramilowski, J. A. et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 1–11 (2015).

95. Ramilowski, J. A. et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 1–12 (2015).

96. Moraes, J. G. N. et al. Uterine influences on conceptus development in fertility-classified animals. *Proc. Natl Acad. Sci. USA* **115**, E1749–E1758 (2018).

97. Cosacak, M. I. et al. Single-cell transcriptomics analyses of neural stem cell heterogeneity and contextual plasticity in a zebrafish brain model of amyloid toxicity. *Cell Rep.* **27**, 1307–1318.e3 (2019).

98. The Igraph Core Team. *igraph – Network analysis software* (The Igraph Core Team, 2019).

99. Hu, H. et al. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38 (2019).

100. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using Tree-based methods. *PLoS ONE* **5**, e12776 (2010).

101. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Author contributions

X.X., H.L., Y.H.L., D.S.C., J.S., Y.H., H.M.Y., W.Q. and L.N.Z. conceived and designed the project. J.C.Z., W.Y.W., H.R.L., F.Y.A., Z.H.C. and Q.H. were responsible for sample collection and dissection. J.C.Z., W.Y.W. and H.R.L. participated in single-nucleus library construction and sequencing. D.S.C., J.C.Z., X.N.D., X.R.W. and T.M.L. coordinated data analysis. L.H.L., R.X., X.L.W., H.Y.W., P.W.D., J.Y.Q., S.Y.W., H.M.L., L.Z., L.H., Y.X.Z., H.X.C., D.M.F., Y.Y.Z., F.Y.W., Y.T.Y., W.D.W. and C.C.S. contributed to single-cell clustering and cell annotation. X.X., H.L., Y.H.L., D.S.C., J.S., Y.H., H.M.Y., W.Q., Y.Q.W., J.H.W., X.H.S., S.J.J., L.C., S.K.S., S.H.Z., X.Y.M., W.M.F., Y.X., F.L., Z.M.L., F.C., X.J., Y.C.X., T.J.W., Y.L., X.M.X., L.L.C., C.C.C., P.L., L.J.Z. and Z.H.O. participated in data interpretation, visualization, and manuscript writing.

Competing interests

The authors declare no competing interests.

Additional information

[1]BGI-Shenzhen, Shenzhen 518083, China. [2]National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Original Bacteria, South China Agricultural University, Guangzhou 510642, China. [3]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China. [4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [5]Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. [6]Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou 510520, China. [7]College of Wildlife Resources Northeast Forestry University, Harbin 150040, China. [8]China National Genebank, BGI-Shenzhen, Shenzhen 518120, China. [9]Shenzhen Key Laboratory of Environmental Microbial Genomics and Application, BGI-Shenzhen, Shenzhen 518120, China. [10]Eye and ENT Hospital, College of Medicine, Fudan University, Shanghai, China. [11]Shanghai Key Laboratory of Visual Impairment and Restoration, Science and Technology Commission of Shanghai Municipality, Shanghai, China. [12]Key Laboratory of Myopia, Ministry of Health, Shanghai, China. [13]HIV and AIDS Malignancy Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-1868, USA. [14]Department of Biochemistry, University of Cambridge, Cambridge CB21QW, UK. [15]Research Center for Translational Medicine, East Hospital, Tongji University School of Medicine, 150 Jimo Road, Shanghai 200120, China. [16]College of Veterinary Medicine, Yangzhou University, Yangzhou 225009, China. [17]Department of Neurology, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou 510080, China. [18]Institute of Special Animal and Plant Sciences (ISAPS) of Chinese Academy of Agricultural Sciences, Changchun, China. [19]Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Shenzhen 518120, China. [20]College of Wildlife and Protected Areas, Northeast Forestry University, No. 26, Hexing Road, Xiangfang District, Harbin 150040, China. [21]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China. [22]Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, 518083 Shenzhen, China. [23]These authors contributed equally: Dongsheng Chen, Jian Sun, Jiacheng Zhu, Xiangning Ding, Tianming Lan, Xiran Wang. ✉email: wildlife530@hotmail.com; gale@scau.edu.cn; liuhuan@genomics.cn; xuxun@genomics.cn