

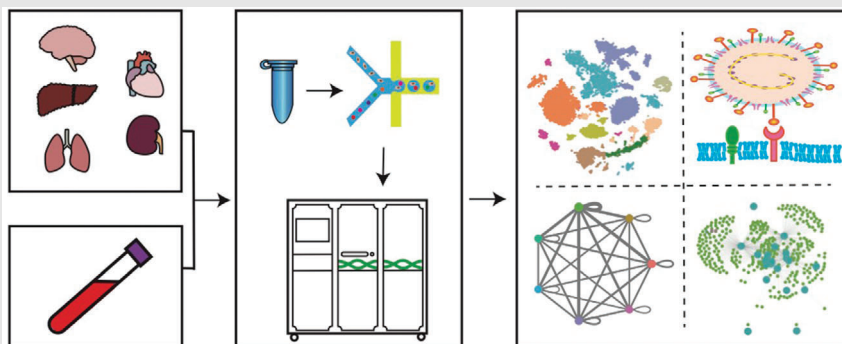
RESEARCH ARTICLE

Screening of cell-virus, cell-cell, gene-gene crosstalk among animal kingdom at single cell resolution

Dongsheng Chen^{1,2}  | Zihua Ou^{1,3} | Jiacheng Zhu^{1,4} | Haoyu Wang^{1,4} |
 Peiwen Ding^{1,4} | Lihua Luo^{1,4}  | Xiangning Ding^{1,4} | Chengcheng Sun^{1,4} |
 Tianming Lan¹ | Sunil Kumar Sahu¹ | Weiyong Wu⁵ | Yuting Yuan⁶ |
 Wendi Wu^{1,7} | Jiaying Qiu^{1,4} | Yixin Zhu^{1,4} | Qizhen Yue^{1,7} | Yi Jia¹ |
 Yanan Wei^{1,7} | Qiuyu Qin^{1,7} | Runchu Li^{1,7} | Wandong Zhao^{1,7} | Zhiyuan Lv^{1,7} |
 Mingyi Pu^{1,7} | Boqiong Lv¹ | Shangchen Yang⁸ | Ashley Chang¹  |
 Xiaofeng Wei⁹ | Fengzhen Chen⁹ | Tao Yang⁹  | Zhenyong Wei⁹ | Fan Yang⁹ |
 Peijing Zhang¹⁰ | Guoji Guo¹⁰ | Yuejiao Li¹ | Yan Hua¹¹ | Huan Liu^{1,4}

¹BGI-Shenzhen, Shenzhen, China²Suzhou Institute of Systems Medicine, Suzhou, Jiangsu, China³Shenzhen Key Laboratory of Unknown Pathogen Identification, BGI-Shenzhen, Shenzhen, China⁴College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China⁵The MOE Frontier Science Center for Brain Research and Brain-Machine Integration, School of Brain Science and Brain Medicine, Zhejiang University, Hangzhou, China⁶Department of Physiology, School of Basic Medical Sciences, Binzhou Medical University, Yantai, China⁷School of Basic Medicine, Qingdao University, Qingdao, China⁸College of Life Sciences, Zhejiang University, Hangzhou, China⁹China National GeneBank, Shenzhen, China¹⁰Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China¹¹Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou, China**Correspondence**

Huan Liu, Yan Hua, Yuejiao Li and Dongsheng Chen.

Email: liuhuan@genomics.cn;wildlife530@hotmail.com;liyuejiao@genomics.cn;dschen2016@gmail.com**Graphical Abstract**

Single-cell atlases of organs and PBMCs in 20 species were constructed. The expression patterns of viral receptors from 24 virus families were characterised in different tissues or PBMC samples to generate a conclusive view on the cellular tropism of viruses. An exploration of cellular communication in brain cells within 11 species was conducted.

RESEARCH ARTICLE

Screening of cell-virus, cell-cell, gene-gene crosstalk among animal kingdom at single cell resolution

Dongsheng Chen^{1,2}  | Zihua Ou^{1,3} | Jiacheng Zhu^{1,4} | Haoyu Wang^{1,4} |
 Peiwen Ding^{1,4} | Lihua Luo^{1,4}  | Xiangning Ding^{1,4} | Chengcheng Sun^{1,4} |
 Tianming Lan¹ | Sunil Kumar Sahu¹ | Weiyang Wu⁵ | Yuting Yuan⁶ |
 Wendi Wu^{1,7} | Jiaying Qiu^{1,4} | Yixin Zhu^{1,4} | Qizhen Yue^{1,7} | Yi Jia¹ |
 Yanan Wei^{1,7} | Qiuyu Qin^{1,7} | Runchu Li^{1,7} | Wandong Zhao^{1,7} | Zhiyuan Lv^{1,7} |
 Mingyi Pu^{1,7} | Boqiong Lv¹ | Shangchen Yang⁸ | Ashley Chang¹  |
 Xiaofeng Wei⁹ | Fengzhen Chen⁹ | Tao Yang⁹  | Zhenyong Wei⁹ | Fan Yang⁹ |
 Peijing Zhang¹⁰ | Guoji Guo¹⁰ | Yuejiao Li¹ | Yan Hua¹¹ | Huan Liu^{1,4}

¹BGI-Shenzhen, Shenzhen, China²Suzhou Institute of Systems Medicine, Suzhou, Jiangsu, China³Shenzhen Key Laboratory of Unknown Pathogen Identification, BGI-Shenzhen, Shenzhen, China⁴College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China⁵The MOE Frontier Science Center for Brain Research and Brain-Machine Integration, School of Brain Science and Brain Medicine, Zhejiang University, Hangzhou, China⁶Department of Physiology, School of Basic Medical Sciences, Binzhou Medical University, Yantai, China⁷School of Basic Medicine, Qingdao University, Qingdao, China⁸College of Life Sciences, Zhejiang University, Hangzhou, China⁹China National GeneBank, Shenzhen, China¹⁰Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou, China¹¹Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou, China**Correspondence**

Huan Liu, Yan Hua, Yuejiao Li and Dongsheng Chen.

Email: liuhuan@genomics.cn;wildlife530@hotmail.com;liyuejiao@genomics.cn;dschen2016@gmail.com**Funding information**

Forestry Science and Technology

Innovation Project of Guangdong

Province, Grant/Award Number:

2022KJCX008; CAMS Innovation Fund

for Medical Sciences(CIFMS),

Grant/Award Number: 22021-I2M-1-061

Abstract

Background: The exact animal origin of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) remains obscure and understanding its host range is vital for preventing interspecies transmission.

Methods: Herein, we applied single-cell sequencing to multiple tissues of 20 species (30 data sets) and integrated them with public resources (45 data sets covering 26 species) to expand the virus receptor distribution investigation. While the binding affinity between virus and receptor is essential for viral infectivity, understanding the receptor distribution could predict the permissive organs and tissues when infection occurs.

Dongsheng Chen, Zihua Ou, Jiacheng Zhu, Haoyu Wang, Peiwen Ding and Lihua Luo contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

Results: Based on the transcriptomic data, the expression profiles of receptor or associated entry factors for viruses capable of causing respiratory, blood, and brain diseases were described in detail. Conserved cellular connectomes and regulomes were also identified, revealing fundamental cell-cell and gene-gene cross-talks from reptiles to humans.

Conclusions: Overall, our study provides a resource of the single-cell atlas of the animal kingdom which could help to identify the potential host range and tissue tropism of viruses and reveal the host-virus co-evolution.

KEYWORDS

crosstalk, single cell sequencing

1 | INTRODUCTION

The coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 has rapidly surged around the world.^{1,2} Understanding the viral host range is essential to prevent cross-species transmissions of zoonotic viruses like SARS-CoV-2 and lots of efforts have been made to predict the potential zoonosis hotspots.^{3–5} Investigations on the host range of viruses using experimental methods can be restricted due to the unavailability of virus/animal/experimental resources, especially for pathogens with high biosafety risks. On the other hand, modelling analysis based on the phylogenetic, ecological and life-history traits of hosts and viruses are able to pinpoint animal species or pathogens that are more prone to cause zoonosis. The combination of experimental evidence and modelling may greatly enhance our preparedness for zoonosis.^{3,6,7} It is known that host cellular surface proteins/ligands are employed by distinct pathogens as their receptors to initiate attachment and penetrate the cell. For example, angiotensin-converting enzyme 2 (ACE2) for severe acute respiratory syndrome coronavirus (SARS-CoV) and SARS-CoV-2,^{8,9} and DPP4 for Middle East respiratory syndrome coronavirus (MERS-CoV).¹⁰ Thus, the viral receptor distribution may reveal the potential replicating niches of the viruses. Transcriptome studies can identify the cellular receptor profiles for the virus of interest, which may be an effective way to narrow down the suspected host and tissue list.¹¹ Identifying the host and tissue tropism is the first step towards understanding viral infection and pathogenesis, thus laying the foundation for the prevention and control of putative outbreaks in animals or humans. scRNAseq data can be utilised to explore conserved or divergent cell-cell and gene-gene cross-talks among multiple species,^{12–15} resulting in various molecular circuits important in development and diseases. In this study, we constructed the single cell atlas for 20 species and characterised the expressions of virus receptors at an unprecedented scale over 75 data sets among 44 animal species. Additionally, we dis-

sected the connectome and regulome of neural cells based on brain single cell transcriptome data of multiple species, thus portraited a comprehensive communication atlas at cellular and genetic level.

2 | RESULTS

2.1 | Generation of single cell atlas for organs and peripheral blood mononuclear cells (PBMCs)

Herein, we constructed the single cell atlas for a total of 20 species (Figure 1, Table S1). The multi-organ atlases were comprised of hedgehog (brain and kidney), alpaca (frontal lobe, liver and lung), mink (frontal lobe, liver and kidney), hamster (frontal lobe, liver, kidney and heart) and chinchilla (brain and kidney) (Figure 1A). The PBMC atlases covered 5 mammal species (horse, tiger, alpaca, red necked wallaby and domestic guinea pig), 1 reptile species (snake) and 10 bird species (dalmatian pelican, black necked crane, red and green macaw, peacock, blue and yellow macaw, helmeted guineafowl, green cheeked parakeet, monk parakeet, sun conure and grey parakeet) (Figure 1B). After quality control (QC), 213 500 and 79 761 cells were maintained for downstream analysis. Cell identity for each cluster was assigned based on the specific expression of canonical cell type markers and the enrichment of GO terms (Tables S1–S3). The brain atlas mainly contained excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, oligodendrocyte progenitor cells, astrocytes and endothelial cells. The kidney atlas was mainly composed of proximal tubule cells, loop of Henle cells, collecting duct principal cells, collecting duct intercalated cells, collecting duct transient cells, podocytes, pericytes, distal convoluted tubule cells, endothelial cells, smooth muscle cells and B cells. The lung atlas included type I alveolar cells (ATI), type II alveolar cells (ATII), ciliated cells, endothelial cells, epithelial cells and fibroblasts. The heart

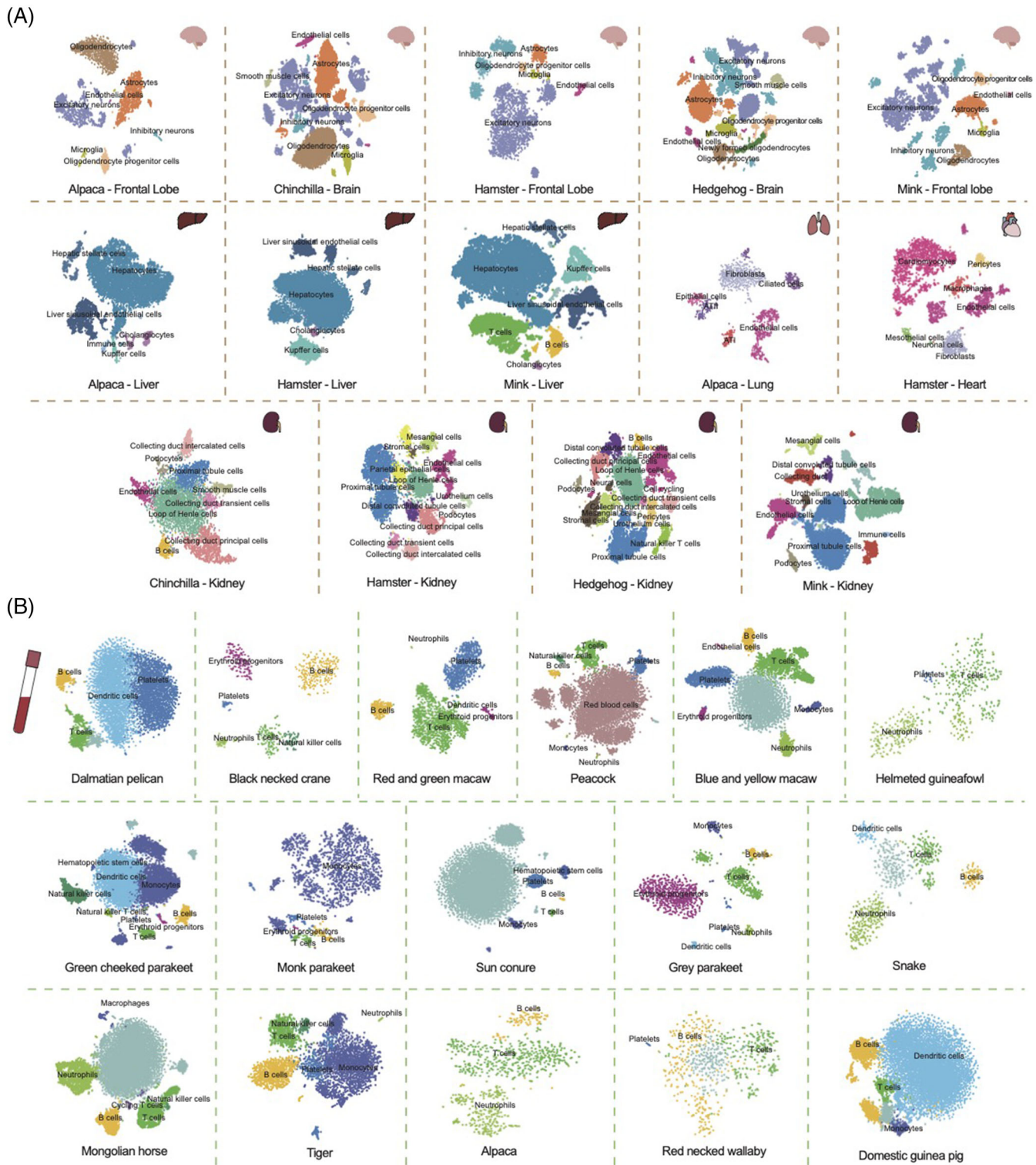


FIGURE 1 Generation of single-cell atlases of 20 species. (A) tSNE plot of hedgehog (brain and kidney), alpaca (frontal lobe, liver and lung), mink (frontal lobe, liver and kidney), hamster (frontal lobe, liver, kidney and heart) and chinchilla (brain and kidney). (B) The PBMC atlases of dalmatian pelican, black necked crane, red and green macaw, peacock, blue and yellow macaw, helmeted guineaowl, green cheeked parakeet, monk parakeet, sun conure, grey parakeet, snake, horse, tiger, alpaca, red necked wallaby and domestic guinea pig.

atlas contained neuronal cells, cardiomyocytes, fibroblasts, mesothelial cells, macrophages, endothelial cells and pericytes. The liver atlas contained liver sinusoidal endothelial cells, cholangiocytes, Kupffer cells, hepatocytes and hepatic stellate cells. The major immune cells, including T cells, B cells, macrophages, natural killer cells and dendritic cells, were identified in PBMC atlas (Figure 1B).

2.2 | Screening of entry factors for coronaviruses in 27 mammalian species

Coronaviruses are among the most common pathogens causing respiratory illnesses in humans.^{16–18} We focused our attention on these viruses because of their roles in epidemics and pandemics. We first screened the expression patterns of several SARS-CoV-2 entry factors and cofactors in the tissues of 27 species (alpaca, mink, chinchilla, domestic guinea pig, hedgehog, horse, red-necked wallaby, human,¹⁹ rhesus monkey,²⁰ crab-eating macaque,^{21,22} pig-tailed macaque [<https://portal.brain-map.org/>], marmoset,²² pig,^{14,23} hamster,¹⁵ mouse,²⁴ blind mole rat,²² rat,^{13,22} rabbit,¹⁵ cat,¹⁵ tiger,¹⁵ civet,²⁵ dog,¹⁵ pangolin,¹⁵ goat,¹⁵ deer,¹⁵ sheep²² and bat²⁶). Angiotensin-converting enzyme 2 (*ACE2*) and the tyrosine-protein kinase receptor UFO (*AXL*) are the cellular receptors responsible for SARS-CoV-2 infection,^{8,27} while trans-membrane serine protease (*TMPRSS2*), Neuropilin-1 (*NRPI*) and the high-density lipoprotein scavenger receptor B type 1 (*SCARBI*) are cofactors promoting the *ACE2*-dependent entry of SARS-CoV-2.^{28–30} Besides, *ACE2* is also a shared receptor for SARS-CoV and HCoV-NL63. Co-expressions of *ACE2* and its cofactors were identified in the tissues of several species (Figure 2, Table S4), including the brain smooth muscle cells and kidney urothelium cells of hedgehog, the kidney distal convoluted tubule cells and liver cholangiocytes of mink, multiple cell types of hamster (the cardiomyocytes of heart, distal convoluted tubule cells, collecting duct transient cells and collecting duct intercalated cells of kidney, sinusoidal endothelial cells, Kupffer cells and hepatocytes of liver, mesothelial cells, ciliated cells, ATI and ATII cells of lung), the kidney B cells of chinchilla, the secretory cells and ciliated cells in the lung of tiger, the ATII cells in the lung of human, the fibroblasts, ATI and ATII cells in the lung of pangolin, the secretory cells and ciliated cells in the lung of bat, the secretory cell and ciliated cells in the lung of cat, ATI cells in the lung of dog, several lung cells of pig (signalling ATII, B cells, ATI and ATII cells), ciliated cells in the lung of mouse, secretory cells, ciliated cells and ATII cells in the lung of rat, ATII and ATI cells in the lung of deer, ciliated cells in the lung of goat. *ACE2* was mostly detected in less than 25% of the cells of a single cell types, which was lower

than *TMPRSS2* and the other receptors for SARS-CoV-2. *AXL*, *NRPI* and *SCARBI* were detected at high levels in the tissues of 21 animals, including hedgehog (brain and kidney), alpaca (frontal lobe, liver and lung), mink (frontal lobe, kidney and liver), hamster (heart, kidney, liver and lung), chinchilla (brain and kidney), tiger (lung and PBMC), human (lung), pangolin (lung), bat (lung), cat (lung), dog (lung), pig (frontal lobe and lung), mouse (lung), rat (lung), blind mole rat (brain), deer (lung), sheep (brain cortex), goat (lung), rabbit (lung), rhesus monkey (dorsolateral prefrontal cortex) and crab-eating macaque (aortic artery, coronary artery and total brain). The close relative of SARS-CoV-2, SARS-CoV, utilises *ACE2*, *CD209*(DC-SIGN), *CLEC4M* (DC-SIGNR) and *CLEC4G* (LSEctin) for cellular entry.^{9,31–33} Among the tissue samples investigated, *CD209* was only detected in the liver of alpaca, the lungs of tiger, human, dog and pig and in the PBMC of domestic guinea pigs. The expression of *CLEC4M* was only observed in a very small fraction of cells in the lungs of human and rat. *CLEC4G* had a higher expression level than *CD209* and *CLEC4M*, especially in alpaca (frontal lobe and liver), hamster (liver), deer (lung) and goat (lung). The receptor for MERS-CoV, *DPP4*, was widely detected in the tissues examined, for example, in the lungs of alpaca, hamster, human, tiger, pangolin, bat, cat, dog, mouse, rat, deer and goat. Aminopeptidase N (ANPEP) was a receptor for HCoV-229E,³⁴ whose expression was found in the lungs of multiple species in our study, including alpaca, hamster, tiger, human, pangolin, bat, cat, dog, pig, mouse, rat and goat. These results showed that the receptors of coronaviruses were widely expressed by mammalian species.

2.3 | Screening of receptors for other respiratory viruses in the lung tissues of 18 animal species

Besides coronaviruses, we also screened for the receptor distribution of other viruses causing respiratory infection. Here, we selected 27 viral receptors corresponding to seven viral families (*Adenoviridae*, *Hantaviridae*, *Paramyxoviridae*, *Parvoviridae*, *Picornaviridae*, *Pneumoviridae* and *Reoviridae*) and screened their expression patterns in the lungs of 18 species (human, pig, mouse, rat, hamster, cat, dog, tiger, civet, rabbit, alpaca, deer, goat, bat, pangolin, duck,¹⁵ pigeon¹⁵ and lizard¹⁵) (Figure S1). *RPSA*, a receptor for a group of parvoviruses including adeno-associated virus 2/3/8/9, was widely expressed by different cell types in the lung of human (Clara cells), mouse (dendritic cells, macrophages, B cells, ATII, endothelial cells, ciliated cells, etc.), rat (monocytes, dendritic cells, macrophages, B cells, T cells, ATI, ATII, fibroblasts, epithelial cells,

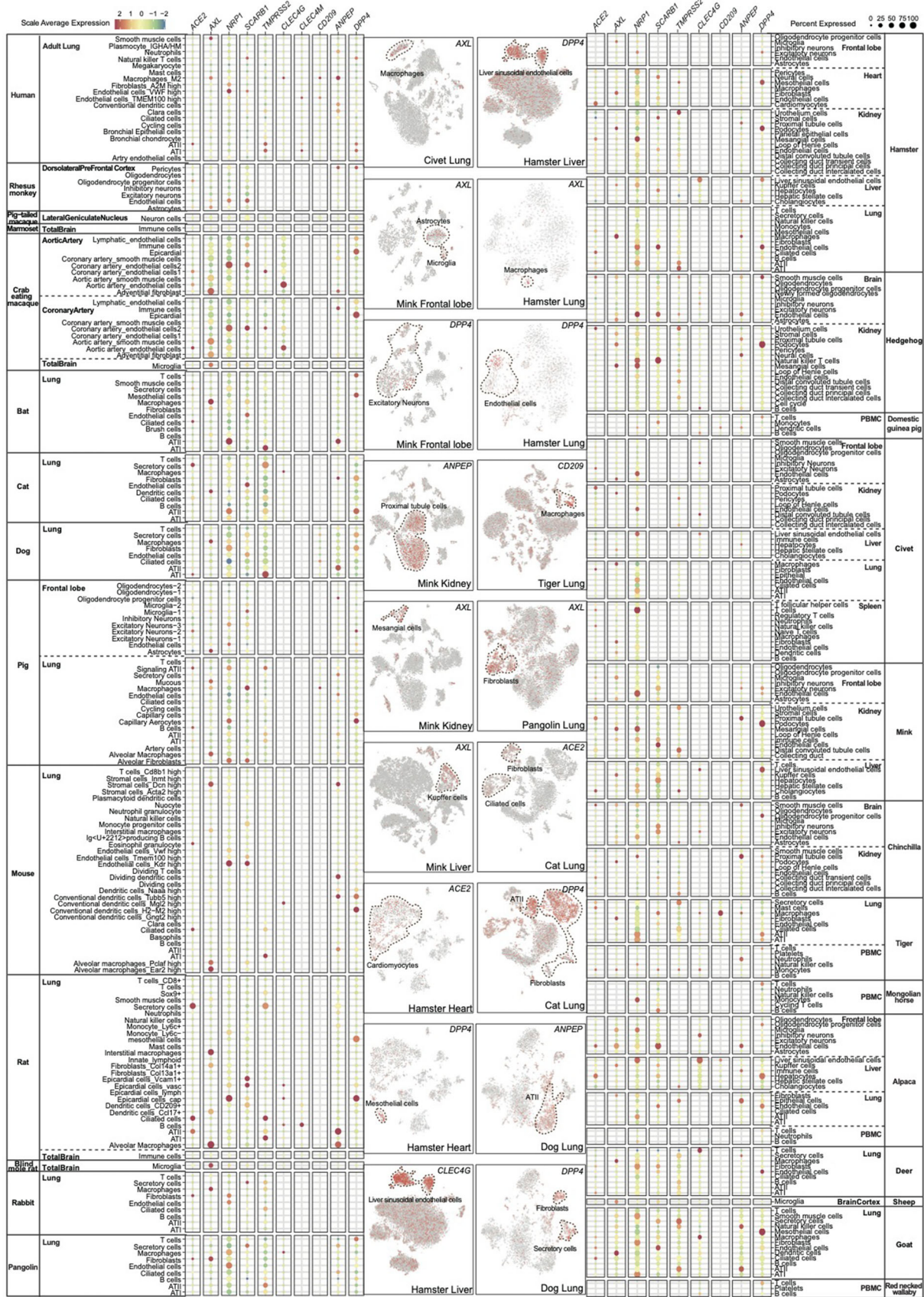


FIGURE 2 Dot plot showing the expression profiles for the entry factors and cofactors of coronaviruses. The expressions of 10 receptors for coronaviruses were screened in tissues of 27 species. Colour saturation reflects the scaled average expression, while dot size indicates the percentage of cells of each cell type expressing the receptor. The feature plots are displayed to show the specific expression of virus receptors in distinct cell types.

ciliated cells, etc.), hamster (secretory cells), tiger (ciliated cells, ATI, ATII, secretory cells, macrophages, endothelial cells, fibroblasts and mast cells) and pigeon (ciliated cells) (Figure S1, Table S4). Besides the above cell types, *RPSA* was also detected by other cells in various species though with lower expression levels and was generally enriched in immune cells such as macrophages, dendritic cells, B cells and T cells. *CDHR3*, the receptor for rhinovirus C of the *Picornaviridae*, was mainly expressed by lung ciliated cells, although sporadic expressions by other cell types were also observed (Figure S1, Table S4). Integrins are a family of transmembrane receptors which could facilitate both cell-cell and cell-extracellular matrix adhesion. They are commonly used as viral receptors for a variety of non-enveloped and enveloped viruses.³⁵ *ITGA5*, a receptor for human parvovirus B19, adeno-associated virus-2 (*Parvoviridae*) and human metapneumovirus (*Pneumoviridae*), was found to be mainly enriched in the endothelial cells of multiple animals including the dog, tiger, cat, bat, pig, civet, alpaca, mouse and lizard (Figure S1). High expression of *ITGA5* was also observed in the macrophages of rat, tiger and rabbit. *ITGAV*, a receptor of human mastadenovirus C (*Adenoviridae*) and human parechovirus 1 (*Picornaviridae*), was commonly expressed by different cell types, including secretory cells, monocytes, mesothelial cells, macrophages, fibroblasts, epithelial, endothelial cells, dendritic cells, B cells, ATII and ATI of different species. In contrast, the expression levels of *ITGB3* (the receptor of hantaan orthohantavirus of *Hantaviridae* and human parechovirus 1 of *Picornaviridae*) were much lower and were mainly observed in the monocytes and innate lymphocytes of rat, and the fibroblasts of dog. *ITGB5*, another receptor of human mastadenovirus C, was commonly observed in fibroblasts (rat, civet, dog, hamster, pangolin and tiger) with some expressions in endothelial cells, ciliated cells, ATI, ATII and macrophages. *CXADR* was a receptor shared by multiple adenoviruses, which was mainly detected in ATI (civet, dog, tiger, cat, rat, mouse, deer, goat, pangolin, rabbit, alpaca and pig), ciliated cells (civet, mouse, rabbit, rat, pangolin, bat, pig and human) and ATII (tiger, cat, deer and pig). *EFNB2* and *EFNB3* are both receptors of Henipaviruses (*Paramyxoviridae*). *EFNB2* was especially enriched in the endothelial cells of 16 species, including deer, goat, duck, civet, tiger, cat, hamster, pigeon, dog, lizard, alpaca, rabbit, pangolin, bat, human and mouse, while *EFNB3* was only sporadically detected in the lung cells of some species. *CD209*, *CD46*, *NECTIN4* and *SLAMF1* are receptors of measles morbilliviruses of the *Paramyxoviridae*. Only *CD46* displayed moderate to high expression levels in the lung cells of some species, such as ciliated cells (hamster, alpaca, rat and bat) and secretory cells (pangolin, hamster and rat).

2.4 | Screening of viral receptors in the PBMC data set of 27 species covering mammals, birds, reptiles and fish

We further explored the putative target cells of viruses capable of causing blood infections (Figure 3, Table S4), based on 46 associated viral entry factors belonging to nine virus families (*Arenaviridae*, *Filoviridae*, *Flaviviridae*, *Hepadnaviridae*, *Herpesviridae*, *Picornaviridae*, *Reoviridae*, *Retroviridae* and *Togaviridae*) using the PBMC data sets of 27 species, which contained the self-produced dataset of 16 species and the public dataset of 11 species (human,³⁶ monkey,³⁷ hamster,³⁸ cat,³⁸ dog,³⁸ mouse,³⁹ rabbit,³⁸ deer,³⁸ goat,³⁸ pigeon³⁸ and zebrafish⁴⁰) (Figure 3). The viruses investigated covered common blood-borne viruses such as human immunodeficiency virus (HIV, *Retroviridae*), hepatitis B virus (HBV, *Hepadnaviridae*) and Hepacivirus C (HCV, *Flaviviridae*), viruses that can cause haemorrhagic fever (*Arenaviridae*, *Filoviridae*, *Flaviviridae* and *Togaviridae*) and others that can cause hepatitis, such as hepatitis A virus (*Picornaviridae*) and hepatitis E virus (*Herpesviridae*). The known receptors for HIVs included *CXCR4*, *CD4CCR5* and *CD209*. While *CCR5* and *CD209* were only sporadically detected in certain species, *CXCR4* was abundantly expressed by B cells of mammals (human, hamster, domestic guinea pig, cat, tiger and rabbit), birds (black necked crane, blue and yellow macaw, dalmatian pelican, green cheek parakeet, grey parakeet, monk parakeet, peacock, pigeon, red and green macaw, sun conure) and snake, and by T cells of at least 13 species (human, alpaca, dog, goat, red necked wallaby, horse, dalmatian pelican, peacock, green cheek parakeet, monk parakeet, pigeon, sun conure and snake). *CD4* was used as a receptor for both HIVs and human betaherpesvirus 7. Interestingly, while *CD4* was found in T cells in most of the studied animals, it was mostly prevalent in dendritic cells and monocytes in humans. Multiple blood cell types displayed low levels of *NPC1* (a receptor of Ebolavirus), including B cells, dendritic cells, macrophages, monocytes, neutrophils, T cells, etc., depending on the species. *AXL*, *TYRO3*, *MERTK*, *CLEC4G*, *CD209* and *HAVCRI* are the main receptors for Ebola virus and Marburg virus. Very low and sporadic expressions of *CLEC4G*, *CD209* and *HAVCRI* were identified in the PBMC data sets. High expression of *AXL* was observed in the dendritic and endothelial cells of rabbit. The arenaviruses shared some of the receptors with filoviruses, including *AXL*, *CLEC4G*, *CD209*, *HAVCRI* and *TYRO3*. Besides these genes, the arenaviruses also use *DAG1*, *LAMP1* and *TFRC* as receptors. *DAG1* exhibited lower expression in different cell types of various animal species, such as B cells (mouse, tiger, red necked wallaby, horse, alpaca, snake and zebrafish), dendritic cells (hamster, domestic guinea pig and snake),



FIGURE 3 Cross-species screening of viral receptor expressions in the PBMC data sets. The expressions of 46 associated viral entry factors in the PBMC of 27 species were screened. Colour saturation of dots reflects the scaled average expression, while dot size indicates the percentage of cells of each cell type expressing the receptor. The corresponding viruses for each entry factor are labelled with different colours and numbers, with virus-specific entry factors on the left side and virus-shared entry factors on the right side.

natural killer cells (hamster, monkey, horse, cat, tiger, black necked crane, green cheek parakeet, zebrafish), platelets (dog, rabbit, grey parakeet, blue and yellow macaw, red and green macaw, black necked crane, sun conure, green cheek parakeet, pigeon) and T cells (horse, red necked wallaby, green cheek parakeet, dalmatian pelican, helmeted guinea fowl). *TFRC* was detected in different cell types in different species, with enrichment in the B cells, dendritic cells, monocytes and plasma cells in the human PBMC data set. HCV (*Flaviviridae*) has several receptors, including *CLDN1*, *EGFR*, *LDLR*, *TFRC*, *EPHA2*, *CD81*, *OCN*, *SCARB1* and *CD209*. Among these, *CD81* displayed higher expressions than the other receptors. For HBV of the *Hepadnaviridae*, the viral receptor gene (*SLC10A1* and *GPC5*) expression was barely detected. For *Herpesviridae*, *CD209*, *CD4*, *CD46*, *ITGA3*, *ITGAV*, *ITGB6*, *ITGB8*, *NECTIN1*, *NECTIN2* and *TNFRSF14* were recognised as receptors. However, only *CD4*, *ITGAV* and *TNFRSF14* displayed sporadic high expression in some species, such as human, hamster and dog. For Enterovirus A71 of the *Picornaviridae*, its receptors (*SCARB2* and *SELPLG*) were distributed in different cell types including T cells, monocytes, dendritic cells and macrophages of different species. Dogs showed higher expression of all the two receptors, while only *SELPLG* had higher expression in human and hamster PBMC. For *Reoviridae*, *HSPA8* was abundantly expressed by B cells, natural killer cells, neutrophil, T cells of 25 species in this study. *ITGA2*, *ITGB1* and *ITGB3* were commonly detected in platelets of mammals and birds, while moderate expression of *FIIR* was only observed in human, hamster, rabbit and tiger. The togaviruses use *MOG*, *MXRA8*, *PHB* and *RPSA* as receptors, and only *RPSA* and *PHB* had high expression in the animal PBMC data set, especially *RPSA*, which also serve as a receptor for flaviviruses.

2.5 | Screening of entry factors for neurotropic viruses in the brain tissues of mammals and reptiles

Neurological complications can be caused by neuroinvasive viruses including polioviruses, rabies viruses, flaviviruses, herpes simplex viruses and measles. In addition to the neuroinvasive viruses, a variety of other viruses such as coronaviruses and influenza viruses, are also capable of infecting the brain tissues under certain circumstances. To explore the putative target cells and hosts of the neurotropic viruses, we systematically screened the expression of 55 viral receptors from 11 viral families (*Arenaviridae*, *Flaviviridae*, *Herpesviridae*, *Paramyxoviridae*, *Phenuiviridae*, *Picornaviridae*, *Pneumoviridae*, *Reoviridae*, *Retroviridae*, *Rhabdoviridae* and *Togaviridae*) in the brain tissues

of 11 species covering 9 mammals (human, monkey, pig, hamster, hedgehog, civet, mink, chinchilla and alpaca) and 2 reptiles (lizard and turtle). *SCARB2*, a receptor for coxsackievirus A16 and enterovirus A71, was highly expressed by oligodendrocytes in the brain (expression percentage > 15%) of human,⁴¹ monkey, hedgehog, civet, mink, alpaca, lizard¹² and turtle¹² (Figure 4, Table S4). In addition to oligodendrocytes, *SCARB2* was also found to be expressed by the endothelial cells of monkey and hedgehog, microglia of chinchilla and mink, and oligodendrocyte progenitor cells of pig and alpaca. *NCAM1* is a receptor for rabies lyssavirus of *Rhabdoviridae*. Conserved expression of *NCAM1* was observed in oligodendrocytes of human, pig, civet, mink, chinchilla and alpaca and oligodendrocyte progenitor cells of all species investigated. Sporadic expressions of *NCAM1* by astrocytes, inhibitory neurons and microglia were also detected. These observations are consistent with a wide host range of rabies viruses. *HSPA8*, a receptor for Simian rotavirus A strain RRV of *Reoviridae*, was widely expressed (expression percentage > 50%) by diverse cell types in the brain of hamster (endothelial cells and inhibitory neurons), mink (inhibitory neurons), chinchilla (oligodendrocytes and microglia), alpaca (endothelial cells, oligodendrocytes and microglia), lizard (excitatory neurons and microglia) and turtle (microglia). *MOG* is a receptor for the rubella virus of *Togaviridae* and was specifically expressed by oligodendrocytes in the brain of mammals including human, monkey, hedgehog, civet, mink, chinchilla and alpaca. *ITGB8*, a receptor for human alphaherpesvirus 1 of *Herpesviridae*, was mainly expressed by oligodendrocytes (hedgehog, civet, mink, chinchilla, alpaca and lizard), oligodendrocyte progenitor cells (human, civet, chinchilla, alpaca and turtle) and astrocytes (hedgehog, civet, mink and chinchilla).

2.6 | Conservation of brain connectomes

In addition to studying the cell-virus interactions, our single-cell resource can be used to investigate the divergence and conservation of cross-talks between cells throughout evolution. To identify putative cellular communications, a ligand-receptor mediated interaction network was constructed for brain cells within 11 species covering 9 mammals (human, monkey, pig, hamster, hedgehog, civet, mink, chinchilla and alpaca) and 2 reptiles (lizard and turtle) (Figure 5A, Table S5). A close interaction between excitatory neurons and astrocytes was observed in human, which was conserved in animals including monkey, hedgehog, mink, chinchilla and alpaca. Whereas the close connection between excitatory neurons and oligodendrocyte progenitor cells was

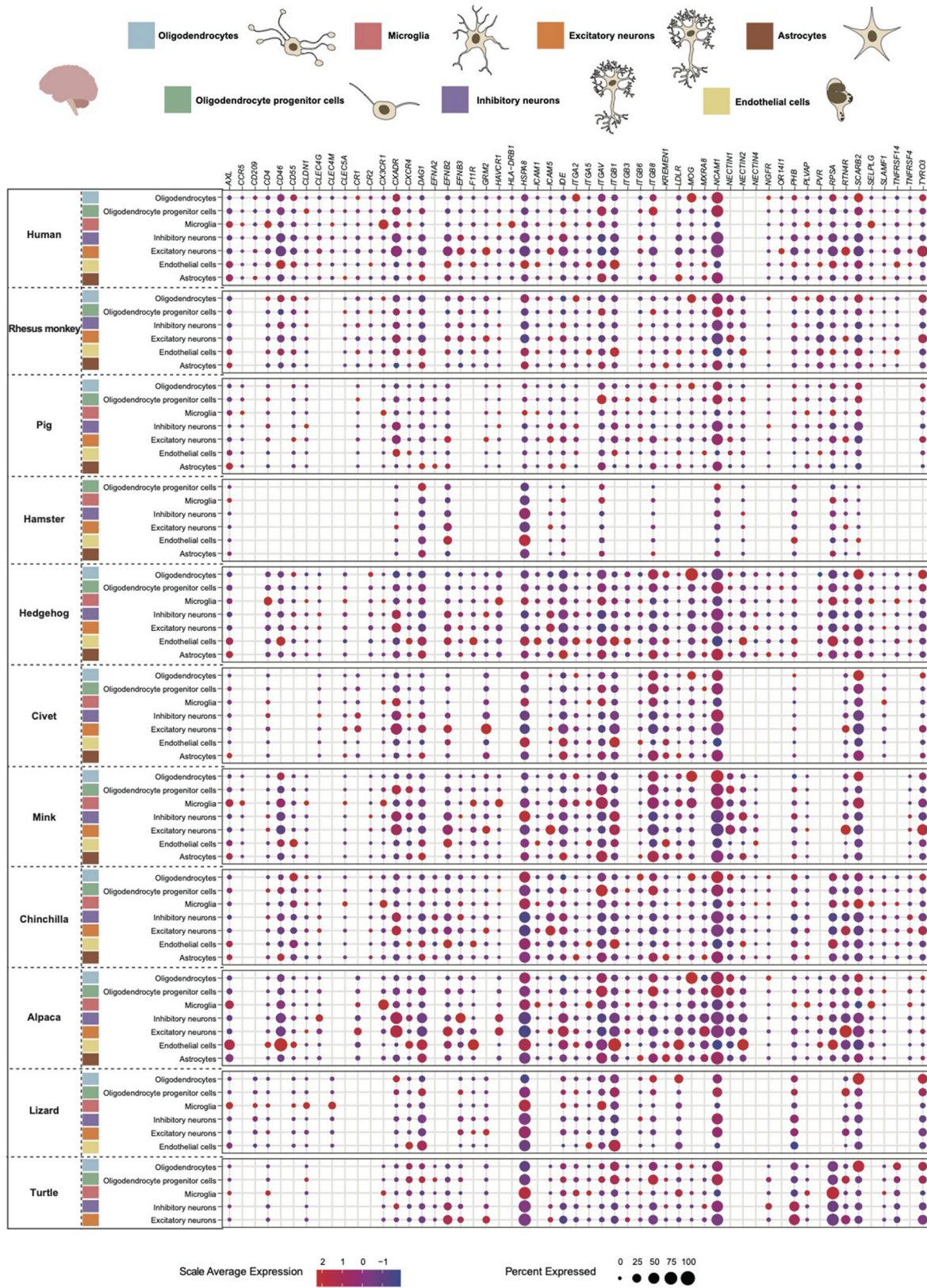


FIGURE 4 Dot plot showing cross-species screening of neurotropic virus putative target cells in brain tissues. The expressions of 55 receptors for neurotropic viruses from 11 viral families (*Arenaviridae*, *Flaviviridae*, *Herpesviridae*, *Paramyxoviridae*, *Phenuiviridae*, *Picornaviridae*, *Pneumoviridae*, *Reoviridae*, *Retroviridae*, *Rhabdoviridae* and *Togaviridae*) in the brain tissues of 11 species covering 9 mammals (human, monkey, pig, hamster, hedgehog, civet, mink, chinchilla and alpaca) and 2 reptiles (lizard and turtle) were screened. Different cell types are annotated with different colours based on scRNA-seq data. Colour saturation of dots reflects the scaled average expression, while dot size indicates the percentage of cells of each cell type expressing the receptor.

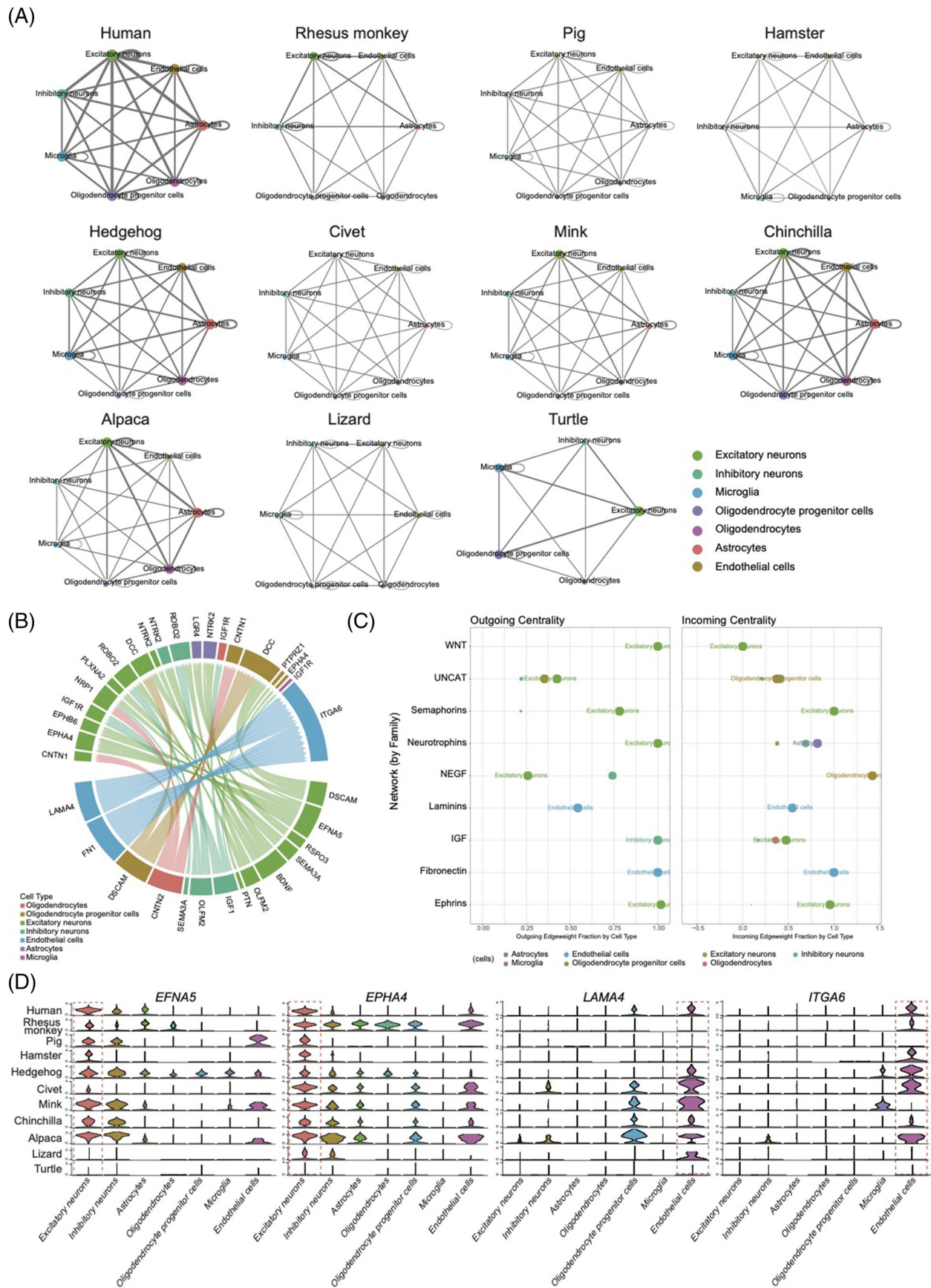


FIGURE 5 Conservation and divergence of brain cell connectomes. (A) A ligand-receptor mediated interaction network for brain cells within 11 species (human, monkey, pig, hamster, hedgehog, civet, mink, chinchilla, alpaca, lizard and turtle). Different cell types are annotated with coloured nodes across species, of which the size is proportional to the sum of ligand-receptor pairs between a cell type with all

specific to human, indicating significant heterogeneity in the cross-talk of brain cells in humans. We next identified pan-conserved cellular connectivity, which may correspond to ancient signalling vectors inherited from common ancestors of mammals and reptiles. In total, we detected 291 pairs of cell-cell interactions conserved among all the 11 species, 44 pairs of which were mediated through the brain derived neurotrophic factor (BDNF)-neurotrophic receptor tyrosine kinase 2 (NTRK2) signalling (Figure 5B, Table S5). The BDNF-NTRK2 signalling plays an important role in neuronal differentiation, maintenance and plasticity.⁴² Here, our results demonstrated that the BDNF-NTRK2 signalling was transduced from excitatory neurons into diverse cell types, including astrocytes, excitatory neurons, inhibitory neurons, microglia and oligodendrocytes. This indicates that excitatory neurons may have a core regulatory role in brain cell connectomes across these 11 species.

In addition to pan-conserved cellular cross-talk, we explored which cell types were predominant in cell-cell cross-talk. In total, we found excitatory neurons appended in most families either outgoing centrality or incoming centrality, and endothelial cells followed closely (Figure 5C). Excitatory neurons received conserved cellular interactions transduced from excitatory and inhibitory neurons, oligodendrocytes progenitor cells and oligodendrocytes, a large portion of which were mediated through inhibitory neurons. Notably, we found that the olfactomedin 2 (OLFM2)-roundabout guidance receptor 2 (ROBO2) signalling could be transduced from inhibitory neurons to excitatory neurons, and vice versa, which has not been observed in neurons before. Intriguingly, we found that the identified pan-conserved cellular cross-talk associated with endothelial cells was entirely mediated between endothelial cells. This indicates that cellular cross-talk between endothelial cells is quite conservative, whereas the interactions between endothelial cells with other types of cells may be species-specific.

The expression of known ligands and receptors was examined to confirm the dominant role of excitatory neurons and endothelial cells across species. For example, ephrin A5 (EFNA5) and EPH receptor A4 (EPHA4) as the ligand and receptor of the ephrin family were commonly expressed in excitatory neurons. The ligand and receptor of laminins family, laminin subunit alpha-4 (LAMA4) and integrin alpha-6 (ITGA6) were also highly expressed

in endothelial cells of most species (Figure 5D). In conclusion, this study systematically revealed both highly conserved and lineage specific cell-to-cell signalling within the vertebrate brain for the first time.

2.7 | Conservation of brain regulomes

To reveal the regulatory mechanisms underlying neural cells from the perspective of evolutionary biology, we predicted the regulomes in neural cells for all the 11 species, resulting in a total of 170 702 TF-target interactions in excitatory neurons and 383 544 TF-target interactions in inhibitory neurons (GENIE3 score > 0.01) conserved in at least 4 species. TF genes and target genes expressed in 11 species ranged from 461 in excitatory neurons cells to 678 in inhibitory neurons cells (Figure 6A and B, Tables S6 and S7). Notably, several well-known regulators for excitatory neurons were active in the genetic regulatory network of the corresponding cell types, consistent with their expected regulatory functions. Additionally, we found a variety of regulatory circuits that were highly conserved among multiple species. Briefly, the interactions between MEF2C its target genes (*EPHA4*, *PEX5L*, *CNTN3* and *PHACTR1*), were conserved in excitatory neurons cells of at least 8 species. In inhibitory neurons, the regulatory relationships between MEF2C and its target genes (*CACNA2D3*, *MMP16*, *PPARGCIA*, *RELN*, *SOX5*, *SOX6*, *ATRNL1* and *KCNC2*) were conserved in more than 8 species. The transcription factor MEF2C is crucial for normal neuronal development in humans and monkey.^{43,44} Consistent with this, our analysis indicated a strongly conserved expression of MEF2C and a conserved core regulatory role of MEF2C across all species.

The regulatory functions of core transcription factors were inferred based on the enriched GO terms of their predicted target genes. The target genes of MEF2C were enriched in GO terms associated with positive regulation of neurogenesis, positive regulation of neuron differentiation and regulation of neuron projection development in either excitatory neurons or inhibitory neurons (Figure 6C and D, Table S8). The conserved and key regulatory roles of MEF2C in both excitatory neurons and inhibitory neurons across species suggest that MEF2C may participate in the maintenance of the balance of cerebral excitatory and inhibitory synapses in a species-conserved

other cell types. The edge thickness is proportional to the sum of ligand-receptor pairs between two cell types. (B) Circos plot displayed the cell-cell interactions conserved among all 11 species. Receptors and ligands occupy the upper and lower semicircle respectively. (C) Network centrality analysis shows the source weight and hub score of various cell types in indicated networks across 11 species. (D) Violin plots showing the conserved expression of *ephrin A5* (*EFNA5*) and *EPH receptor A4* (*EPHA4*) in excitatory neurons, and laminin subunit alpha-4 (*LAMA4*) and integrin alpha-6 (*ITGA6*) in endothelial cells across 11 species.

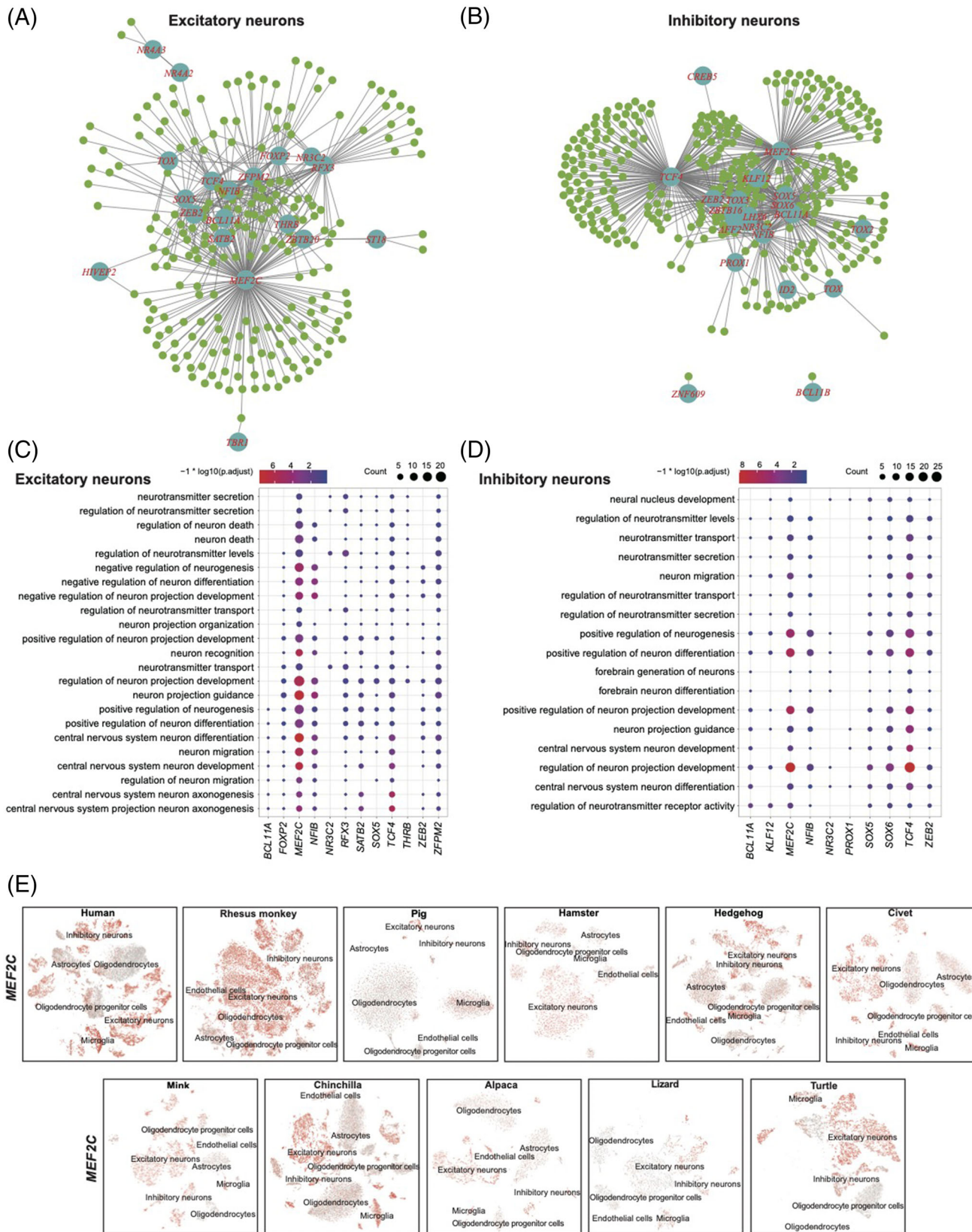


FIGURE 6 Conservation and divergence of brain cell regulomes. Gene regulatory network of the conserved TFs and target genes detected across 11 species in excitatory neurons (A) and inhibitory neurons (B). Blue nodes with labels represent the TFs and the smaller green nodes represent the target genes. GO enrichment of the target genes of key TFs in excitatory neurons (C) and inhibitory neurons (D). Colour saturation of dots reflects the significance level of enrichment analysis, while dot size indicates the count of target genes classified in each GO term. (E) Feature plots displaying the conserved expression of MEF2C in excitatory neurons and inhibitory neurons across 11 species.

manner. Whereas the different conserved target genes of MEF2C in excitatory neurons and inhibitory neurons indicate the diverse regulatory functions of MEF2C in different cell types. We also found that MEF2C was commonly expressed in all the studied species (Figure 6E). Overall, our study systematically identified a conserved core regulatory program for brain cells, including both well-recognised and novel regulators (Figure 6).

3 | DISCUSSION

Overall, this study characterised the expression of receptors used by a variety of viruses in 75 tissues from 44 species from mammals, birds, reptiles and fishes. The resources provided by this study may assist the exploration of tissue and host tropism of corresponding viruses, direct experimental design and guide the choice of appropriate animal infection models.

Defining the pathogen's key cellular receptors is the first step in determining its potential host range. With the key binding sites unravelled, genomic data can be used to compare the structural similarity of the receptor homologs between different species. If the structural similarity is high enough to support potential interaction between the virus and an animal species, we can further move to construct the protein sequences of the viral binding domains and the receptor homologs, then assess their binding affinity using experimental methods such as pseudovirus infection assay, flow cytometry, western blot and ELISA. With proper positive and negative control, we would be able to know the receptor of which species may support virus entry. However, different viruses vary in their infection routes and tissue tropisms, a further understanding of the viral receptor distribution and expression levels in different organs would help reveal the anatomical sites potentially infected or diseased by the virus. Such analysis can be readily supported by the scRNA-seq data as provided in this study. Further confirmation studies using *in vitro* and/or *in vivo* infection experiments, or field surveillance data originated from serology test or nucleotide detection can finally confirm the susceptibility of a certain animal species to our target pathogen. Following this rationale, we can narrow down the suspect host list step by step (Figure 7). From the very beginning of the COVID-19 pandemic, scientists have already tried to pinpoint the animal host range of SARS-CoV-2 using some of the above methods.^{8,45–52} Upon the discovery of the etiological agent of COVID-19, ACE2 has been recognised as the receptor of SARS-CoV-2, and the critical sites composing the binding surface between the viral spike protein and ACE2 was soon identified.^{8,45,46} Later, a comparative study on the conservation of ACE2 in different animal species

was carried out using the ACE2 sequence data.⁴⁷ The binding affinities between a bunch of ACE2 orthologs and the functional domains of SARS-CoV-2 spike protein were also evaluated.⁴⁸ To reveal the tissue tropism of SARS-CoV-2 in the human body, the expression landscape of ACE2 and its cofactor TMPRSS2 was characterised by several groups.^{49–52} Similarly, our group has utilised scRNA-seq data to explore the host range and tissue tropism of SARS-CoV-2.¹⁵ It's noteworthy that when ACE2 in white-tailed deer was reported to exhibit high binding scores to SARS-CoV-2 spike protein,⁴⁷ subsequent serology surveillance in wild individuals and laboratory infection experiments further supported this species as a permissive host of SARS-CoV-2,^{53,54} extending our understanding of SARS-CoV-2 behaviour in the real-world. However, it should be noted that most of the related research work was individually conducted by different groups separately and perhaps simultaneously, which didn't help to reduce the suspect list in a stepwise manner and each step remains resource expensive. What we called for, is the close interdisciplinary collaborations between bioinformatic and bench scientists to realise the OneHealth concept, through effective integration of multi-omics data including genomic, transcriptomic, proteomic and surveillance data. The related technologies should be progressively combined and operated within an organised network, so that our scientific questions can be solved more systematically, effectively and economically. Despite the positive outlook, we must recognise that much work remained to be done to fulfil this framework. First, the genome sequences of a lot of species are still lacking. For example, bats are reservoir hosts for abundant viruses, yet less than 100 of the over 1400 bat species have their genomes sequenced^{55–57} (<https://bat1k.com/progress/>), hindering our investigation on the role of these animals in virus maintenance and transmission. Second, a virus may employ not just one cellular surface protein as its binding target, and the involvement of alternative receptors inevitably complicates our judgement on viral infectivity and pathogenesis. Built on previous experience of SARS-CoV, Zhou et al.^{8,30} soon confirmed ACE2 as an entry factor for SARS-CoV-2, with TMPRSS2 as its cofactor. However, the common pneumonia cases caused by SARS-CoV-2 seemed to contradict the low co-expression of ACE2 and TMPRSS2 in human lung cells, as revealed in this and previous reports.^{49,58} Later, other receptors and entry cofactors of SARS-CoV-2 were uncovered, including *AXL*, *NRPI*, *SCARB1* etc.,^{27–29} which may explain its infection in tissues with low ACE2 expression. This kind of hidden information may be responsible for the incongruence between binding affinity results and field observations. However, it remains hard to use scRNA-seq data to measure how the different entry factors and cofactors contribute to the *in vivo* viral infection profile,

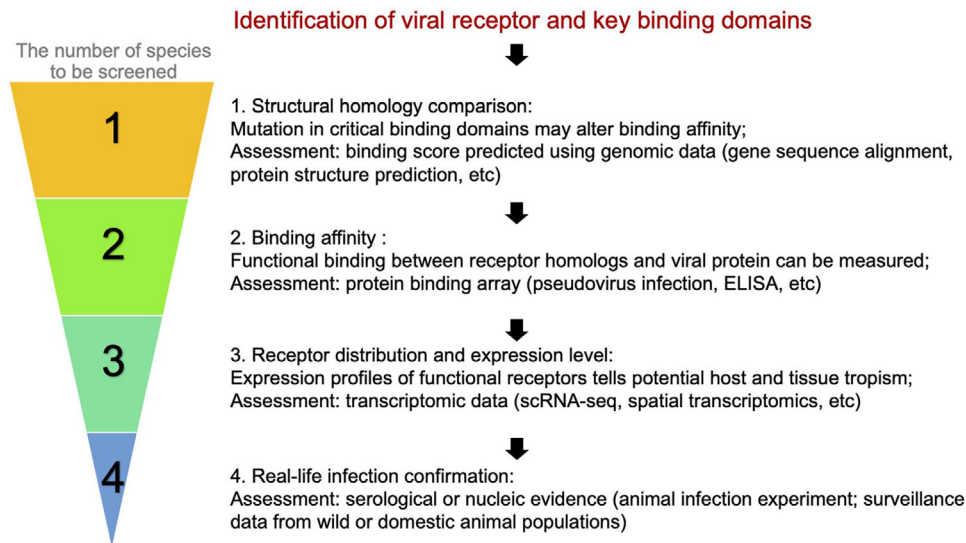


FIGURE 7 An integrative multi-omics workflow to identify the host range of novel pathogens.

such as the expression threshold for a positive prediction. Furthermore, little is known about the entry factors of many viruses. Third, not all viruses use proteins as their receptors. A prominent example will be influenza A viruses, which utilise sialic acid with different linkages as receptors.⁵⁹ Such post-translational modifications can't be predicted based on transcriptomic data. Fourth, receptor binding affinity is the mutual effect of viral and host protein. The binding affinity between a defined virus variant and a receptor protein can't fully reflect the interaction behaviour between a new variant and the same receptor. Therefore, the virus binding affinity must be determined separately for different species and even variants within the same species. Fifth, transcriptomic data for many species and tissues remain scarce. Such data may be available for human and classical animal models but not for the thousands of wild species involved in the field of virology, although network centrality studies suggest carnivores, bats and ungulates hold central positions in sharing viruses with other hosts.^{3,4} The process to generate their transcriptome data, especially scRNA-seq data, would be time and resource consuming. However, with the advancement of technology and accumulation of genomic and transcriptomic data, we should be aware of the potential benefits that these massive multi-omics data could bring to virology study and build up the knowledge system with an appropriate design. Right now, the research rationale we proposed may appear implausible and distant, but one day, scientists will be able to quickly predict the potential host and tissue tropism of novel zoonotic pathogens based on these resources before turning to animal infection experiments and field investigation.

To have a conclusive view of the cell tropism of viruses, we combined all the receptor expressions obtained by this study and used a threshold of > 15% expression and > 0.5 scaled expression level to filter out receptors with obvious expressions. Nine receptors (CDHR3, CXADR, CXCR4, EFNB2, HSPA8, ITGA2, ITGB1, ITGB3 and RPSA) displayed cell-specific expression in at least 10 species. In the lung, CDHR3 was enriched in the ciliated cells in 15 species, serving as a receptor of rhinovirus C viruses, indicating that ciliated cells in the respiratory tract may be the main target of this virus. CXADR was a receptor for human mastadenoviruses and was mainly expressed by lung ATI cells. EFNB2 was mainly expressed by lung endothelial cells, facilitating the cell entry of Nipah henipavirus. In the blood, B cells, T cells and platelets support the expression of multiple viral receptors. CXCR4, the receptor for HIV-1, HIV-2 and feline immunodeficiency virus, was highly expressed in both B and T cells. A similar distribution was also identified for RPSA, which helps the cell binding for various viruses, including classical swine fever virus, dengue virus, Sindbis virus and Venezuelan equine encephalitis virus. HSPA8, a receptor of simian rotavirus A strain RRV, was highly expressed by the T cells of 14 species. The platelets commonly expressed three receptors, including ITGA2 (a receptor for human rotavirus strain WA and simian rotavirus A strain RRV), ITGB1 (a receptor for human rotavirus strain WA, mammalian orthoreovirus and simian rotavirus A strain RRV) and ITGB3 (a receptor for West Nile virus and simian rotavirus A strain RRV). Understanding the distribution of commonly preferred receptors could help explain the potential tissue tropisms of multiple pathogens. The cells expressing receptors may have a higher chance to assist

the initiation of viral infection and promote the intra-host spread of the viruses. Drugs targeting common major viral receptors may facilitate the prevention and treatment of multiple virus infections.

However, there are several limitations in our study and the results should be interpreted cautiously. (i) We were only able to obtain scRNAseq data for limited number of tissues from each animal species in this study, due to difficulty in sampling and experimental failures. A more comprehensive characterisation of the disease-related organs is needed for the interpretation of clinical symptoms induced by infections of various viruses in the animals. (ii) The dataset of each organ for each species lacked biological and technical replicates. Therefore, we failed to reveal the comprehensive cell type compositions of each tissue, which hindered the cross-species comparisons on cell types. Our data may also be influenced by sampling bias due to the individual differences in age, gender, health status, etc. To reduce sampling bias, we have tried to use a relatively high filtering threshold to keep only reliable results. (iii) The expression profiles of the receptors were characterised based on the RNA sequencing data, rather than their mature protein form with appropriate post-translational modifications. The influenza A viruses, as mentioned above, utilise sialic acid receptors and the sialyl linkage types to determine the differential binding specificity of avian and human adapted strains.⁵⁹ Although the transcriptome data could reveal the presence of the proteins associated with the receptors, the exact expression profiles of their receptors could not be stated by this data. Therefore, influenza viruses were not included in this study. (iv) We have only characterised the known receptor genes, other potential alternative receptors of the viruses were not included, but their distribution could also affect the tissue tropism of the viruses. (v) Both the expression profiles and the binding affinity between key cellular receptors and viral proteins are essential for viral entry into the host cells. Because protein products of the same ortholog may have different structures, and mutations in the binding domain may reduce or even abolish its binding ability to the corresponding viral surface proteins. The tissue tropism and host range of the viruses should be confirmed by experiments or epidemiological evidence. (vi) For some animal species lacking high-resolution genomic data, accurate determination of their gene expression levels remains challenging. The viral receptor screening, connectome and regulome analyses are based on bioinformatic predictions. Therefore, our data needs to be interpreted with caution and awaits further experimental validation.

In addition to investigating the cell-virus interactions, the datasets generated in this study can provide a comprehensive data source for the deep analysis of cross-talks

between cells and genes. Cell-cell communication is the main driving factor of cell differentiation and physiological function, which determines tissue/organ development, homeostasis and response to injury. In tissues, cells communicate directly with local neighbours through paracrine signalling and cell-cell direct contact, or exchange information with remote partners through endocrine signalling. The emergence of single-cell technology makes it possible to reveal cell-specific ligand receptor patterns comprehensively. Here, we explored the cellular communications for brain cells within 11 species, and identified the excitatory neurons as the predominant cell type in cell-cell cross-talk. The investigation of the regulomes in neural cells for all the 11 species also identified conserved regulators for brain cells. These results suggest that the generated datasets can facilitate the research community for their personalised needs.

4 | MATERIAL AND METHODS

4.1 | Ethics statement

This project was approved by the Institutional Review Board on the Ethics Committee of BGI (Approval letter reference number BGI-NO. BGI-IRB A20008-T2). All procedures were conducted according to the guidelines of the Institutional Review Board on the Ethics Committee of BGI.

4.2 | Materials description

The construction of transcriptome libraries was first performed on 20 species, of which 5 species, alpaca (*Vicugna pacos*), mink (*Neovison vison*), chinchilla (*Chinchilla lanigera*), hedgehog (*Atelerix albiventris*) and hamster (*Mesocricetus auratus*), were performed using 10× Genomics' Chromium and single cell sequencing libraries. As for green cheeked parakeet (*Pyrrhura molinae*), red and green macaw (*Ara chloropterus*), grey parakeet (*Psittacus erithacus*), dalmatian pelican (*Pelecanus crispus*), peacock (*Pavo cristatus*), tiger (*Panthera tigris altaica*), domestic guinea pig (*Cavia porcellus*), alpaca (*Vicugna pacos*), monk parakeet (*Myiopsitta monachus*), helmeted guineafowl (*Numida meleagris*), snake (*Ptyas mucosa*), blue and yellow macaw (*Ara ararauna*), sun conure (*Aratinga solstitialis*), horse (*Equus caballus*), red necked wallaby (*Notamacropus rufogriseus banksianus*) and black necked crane (*Grus nigricollis*), were performed using scRNA-seq with DNBelab C4 platform. Next, BGI-seq 500 sequencer was used for the sequencing of transcriptome libraries. In addition to the self-generated data, we

also retrieved single-cell transcriptome sequencing data from 26 species, civet (*Paguma larvata*), bat (*Rhinolophus sinicus*), blind mole rat (*Nannospalax galili*), cat (*Felis catus*), crab-eating macaque (*Macaca fascicularis*), deer (*Cervus nippon*), dog (*Canis lupus familiaris*), goat (*Capra aegagrus hircus*), hamster (*Mesocricetus auratus*), human (*Homo sapiens*), marmoset (*Callithrix jacchus*), mouse (*Mus musculus*), pangolin (*Manis javanica*), pig (*Sus scrofa domestica*), pig-tailed macaque (*Macaca nemestrina*), rabbit (*Oryctolagus cuniculus*), rat (*Rattus norvegicus*), rhesus monkey (*Macaca mulatta*), sheep (*Ovis aries*), tiger (*Panthera tigris altaica*), pigeon (*Columba livia domestica*), zebrafish (*Danio rerio*), lizard (*Pogona vitticeps*, *Anolis carolinensis*), turtle (*Trachemys scripta elegans*), duck (*Anas platyrhynchos domesticus*), from the public papers or databases (Table S1).

4.3 | Sample treatment

Tissue samples were dissected and rinsed with 1× PBS to remove the surface blood. Next, the processed samples were frozen in liquid nitrogen for about 10 minutes, and then transferred to a liquid nitrogen tank for storage till further use. The whole blood sample of the animal was taken from the adult living animal and then immediately placed in an EDTA anticoagulant tube. Tissue samples and PBMC samples were treated following previous protocols.¹⁵

4.4 | Homologs conversion

Homologs were downloaded from BioMart and genes from each species were converted to human homologs to allow cross-species comparison. In situation of species lacking homologs in Ensembl database, we identified single-copy orthologs using OrthoFinder v2.3.3 (orthofinder -f input_dir -t 10 -a 10).⁶⁰ The analysis is to perform sequence alignment for the protein FASTA files of two species, and subsequently calculate the orthologous relationship of protein-coding genes between different species. If a 1:1 match existed between a non-human and human gene, the non-human gene name was converted to the human gene name. And other genes that are not a one-to-one match will be filtered out.

4.5 | Single-nucleus RNA-sequencing data processing

Single-nucleus RNA-sequencing (snRNA-seq) data and gene expression matrix were obtained using Cell Ranger

3.0.2 (10× Genomics) and all gene expression matrices were processed with R package Seurat.⁶¹ Briefly, cells with gene numbers lower than 200 were filtered out. Cells with mitochondrial percentage higher than 10% were discarded as well. Then, ‘LogNormalize’ method was used for data normalisation. ‘FindVariableFeatures’ function implemented in Seurat was utilised to select the top 2000 highly variable genes. Subsequently, data were scaled and principal component analysis (PCA) was conducted using variable genes. Significant principal components (p -value less than .01) were selected for clustering with the shared nearest neighbour (SNN) based method implemented in ‘FindClusters’ with resolution of 1.0. Finally, the clustering result was visualised using t-distributed stochastic neighbour embedding (t-SNE) graphs.

4.6 | Differential expression analysis and cell types annotation

Differentially expressed genes (DEGs) were identified by ‘FindAllMarkers’ function in Seurat. p Values were adjusted with Benjaminiand Hochberg (BH) method, and pathways with an adjusted p value $\leq .05$ were considered significantly enriched. DEGs, GO pathways and verified markers were combined to annotate cell types for each data set. First, canonic cell type markers were collected from published literatures. The expression patterns of cell type marker were used to infer cell type identity. DEG gene list was used to perform GO term enrichment analysis using clusterProfiler package⁶² in order to confirm the cell type annotation results.

4.7 | Public data collection and processing

Public single cell data sets were downloaded from NCBI, GEO, Human Cell Atlas Portal and EBI single cell database. All public data sets were processed using the same criteria according to the previous studies.

4.8 | Prediction of host tropism and target cells of viruses

All viruses and corresponding receptors were collected from the viral receptor database (<http://www.computationalbiology.cn:5000/viralReceptor>),⁶³ Human Lung Cell Atlas (HLCA) database⁶⁴ or published articles.^{30,65,66} The expression pattern of these receptors was calculated in all cell types of each data set.

4.9 | Brain cellular communication analysis

Brain data sets from 11 species (*Homo sapiens*, *Macaca mulatta*, *Sus scrofa domestica*, *Mesocricetus auratus*, *Eriaceus europaeus*, *Paguma larvata*, *Neovison vison*, *Chinchilla lanigera*, *Vicugna pacos*, *Pogona vitticeps* and *Trachemys scripta elegans*) were used to perform communication analysis using R package Connectome. We selected the FANTOM5 database⁶⁷ which included all literature-supported human ligands and receptors. Connectome networks were constructed according to 'CreateConnectome' function. We selected conserved cellular interactions in brain data sets from these 11 species based on the following criteria: ligands and receptors expressed in all data sets and the interaction expressed in at least 8 data sets. A Circos plot was used to visualise conserved interactions.

4.10 | Brain TF-target interaction analysis

The human TF list was downloaded from the animalTFDB,⁶⁸ and protein coding genes were selected for downstream analysis. We also filtered out the TFs and genes which were expressed in less than 5% of excitatory and inhibitory neurons. For each data set, putative regulatory circuits were predicted by GENIE3 package,⁶⁹ and TF-target interactions with weight value ≥ 0.01 were retained.


ACKNOWLEDGEMENTS

This study was supported by Forestry Science and Technology Innovation Project of Guangdong Province (2022KJCX008) and CAMS Innovation Fund for Medical Sciences (CIFMS)(22021-I2M-1-061). Our work was supported by the China National GeneBank (CNGB), and we thank the CNGB for producing and storing the sequencing data (CNP0002081).

CONFLICT OF INTEREST

The authors declare no competing interests.

ORCID

Dongsheng Chen  <https://orcid.org/0000-0003-2400-0315>

Lihua Luo  <https://orcid.org/0000-0002-2129-4226>

Ashley Chang  <https://orcid.org/0000-0002-3439-9457>

Tao Yang  <https://orcid.org/0000-0003-4211-5183>

REFERENCES

1. Wu F, Zhao Su, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265–269.
2. Zhu Na, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727–733.
3. Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. *Nature*. 2017;546:646–650.
4. Wells K, Morand S, Wardeh M, Baylis M. Distinct spread of DNA and RNA viruses among mammals amid prominent role of domestic species. *Glob Ecol Biogeogr*. 2020;29:470–481.
5. Allen T, Murray KA, Zambrana-Torrel C, et al. Global hotspots and correlates of emerging zoonotic diseases. *Nat Commun*. 2017;8:1–10.
6. Gibb R, Redding DW, Chin KQ, et al. Zoonotic host diversity increases in human-dominated ecosystems. *Nature*. 2020;584:398–402.
7. Albery GF, Becker DJ, Brierley L, et al. The science of the host-virus network. *Nat Microbiol*. 2021;6:1483–1492.
8. Peng Z, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270–273.
9. Li W, Moore MJ, Vasileva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*. 2003;426:450–454.
10. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*. 2015;1282:1–23.
11. Chen D, Tan C, Ding P, et al. VThunter: a database for single-cell screening of virus target cells in the animal kingdom. *Nucleic Acids Res*. 2022;50:D934–D942.
12. Tosches MA, Yamawaki TM, Naumann RK, Jacobi AA, Tushev G, Laurent G. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science*. 2018;360:881–888.
13. Raredon MSB, Adams TS, Suhail Y, et al. Single-cell connectomic analysis of adult mammalian lungs. *Sci Adv*. 2019;5(12):eaaw3851.
14. Zhu J, Chen F, Luo L, et al. Single-cell atlas of domestic pig cerebral cortex and hypothalamus. *Sci Bull*. 2021;66:1448–1461.
15. Chen D, Sun J, Zhu J, et al. Single cell atlas for 11 non-model mammals, reptiles and birds. *Nat Commun*. 2021;12.
16. Zhang S-F, Tuo J-L, Huang X-B, et al. Epidemiology characteristics of human coronaviruses in patients with respiratory infection symptoms and phylogenetic analysis of HCoV-OC43 during 2010–2015 in Guangzhou. *PLoS One*. 2018;13:1–20.
17. To KKW, Chan K-H, Ho J, et al. Respiratory virus infection among hospitalized adult patients with or without clinically apparent respiratory infection: a prospective cohort study. *Clin Microbiol Infect*. 2019;25:1539–1545.
18. Nickbakhsh S, Ho A, Marques DFP, Mcmenamin J, Gunson RN, Murcia PR. Epidemiology of seasonal coronaviruses: establishing the context for the emergence of coronavirus disease 2019. *J Infect Dis*. 2020;222:17–25.
19. Han X, Zhou Z, Fei L, et al. Construction of a human cell landscape at single-cell level. *Nature*. 2020;581:303–309.
20. Zhu Y, Sousa AMM, Gao T, et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science (80-)*. 2018;362.
21. Zhang W, Zhang S, Yan P, et al. A single-cell transcriptomic landscape of primate arterial aging. *Nat Commun*. 2020;11.

22. Geirsdottir L, David E, Keren-Shaul H, et al. Cross-species single-cell analysis reveals divergence of the primate microglia program. *Cell*. 2019;179:1609–1622.e16.
23. Zhang L, Zhu J, Wang H, 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*. 2021;48:411–425.
24. Han X, Wang R, Zhou Y, et al. Mapping the mouse cell atlas by Microwell-Seq. *Cell* 2018;172:1091–1107.e17.
25. Ding P, Wang H, Zhu J, et al. Viral receptor profiles of masked palm civet revealed by single-cell transcriptomics [published online ahead of print, 2022 Apr 28]. *J Genet Genomics*. 2022;S1673-8527(22)00129-1
26. Wan X, Ou Z, Ding P, et al. Comparative analysis of single cell lung atlas of bat, cat, tiger and pangolin. 2021; *bioRxiv*. 2021.12.26.473325. <https://doi.org/10.1101/2021.12.26.473325>
27. Wang S, Qiu Z, Hou Y, et al. AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells. *Cell Res*. 2021;31:126–140.
28. Wei C, Wan L, Yan Q, et al. HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry. *Nat Metab*. 2020;2:1391–1400.
29. Cantuti-Castelvetri L, Ojha R, Pedro LD, et al. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science (80-)*. 2020;370:856–860.
30. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020;181:271–280.e8.e8.
31. Yang Z-Y, Huang Y, Ganesh L, et al. pH-dependent entry of severe acute respiratory syndrome coronavirus is mediated by the spike glycoprotein and enhanced by dendritic cell transfer through DC-SIGN. *J Virol*. 2004;78:5642–5650.
32. Marzi A, Gramberg T, Simmons G, et al. DC-SIGN and DC-SIGNR interact with the glycoprotein of Marburg virus and the S protein of severe acute respiratory syndrome coronavirus. *J Virol*. 2004;78:12090–12095.
33. Gramberg T, Hofmann H, Möller P, et al. LSEctin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology*. 2005;340:224–236.
34. Yeager CL, Ashmun RA, Williams RK, et al. Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature*. 1992;357:420–422.
35. Stewart PL, Nemerow GR. Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol*. 2007;15:500–507.
36. Zhu L, Yang P, Zhao Y, et al. Single-cell sequencing of peripheral mononuclear cells reveals distinct immune response landscapes of COVID-19 and influenza patients. *Immunity* 2020;53:685–696.e3.
37. Kotliar D, Lin AE, Logue J, et al. Single-cell profiling of ebola virus disease in vivo reveals viral and host dynamics. *Cell* 2020;183:1383–1401.e19.
38. Li Z, Sun C, Wang F, et al. (2022). Molecular mechanisms governing circulating immune cell heterogeneity across different species revealed by single-cell sequencing. *Clinical and Translational Medicine*, 12(1). Portico. <https://doi.org/10.1002/ctm2.689>
39. Winkler ES, Shrihari S, Hykes BL, et al. The intestinal microbiome restricts alphavirus infection and dissemination through a bile acid-type I IFN signaling axis. *Cell* 2020;182:901–918.e18.
40. Tang Q, Iyer S, Lobbardi R, et al. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med*. 2017;214:2875–2887.
41. Velmeshev D, Schirmer L, Jung D, et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* 2019;364:685–689.
42. Oh H, Lewis DA, Sibille E. The role of BDNF in age-dependent changes of excitatory and inhibitory synaptic markers in the human prefrontal cortex. *Neuropsychopharmacology* 2016;41:3080–3091.
43. Meur NL, Holder-Espinasse M, Jaillard S, et al. MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J Med Genet*. 2010;47:22–29.
44. Yelamanchili SV, Chaudhuri AD, Chen LN, Xiong H, Fox HS. MicroRNA-21 dysregulates the expression of MEF2C in neurons in monkey and human SIV/HIV neurological disease. *Cell Death Dis*. 2010;1.
45. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;581:215–220.
46. Wang Q, Zhang Y, Wu L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell* 2020;181:894–904.e9.
47. Damas J, Hughes GM, Keough KC, et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc Natl Acad Sci USA*. 2020;117:22311–22322.
48. Wu L, Chen Q, Liu K, et al. Broad host range of SARS-CoV-2 and the molecular basis for SARS-CoV-2 binding to cat ACE2. *Cell Discov*. 2020;6:1–12.
49. Sungnak W, Huang N, Bécavin C, et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med*. 2020;26:681–687.
50. Xu H, Zhong L, Deng J, et al. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci*. 2020;12:1–5.
51. Zou X, Chen K, Zou J, et al. Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front Med*. 2020;14:185–192.
52. Wang Z, Xu X. scRNA-seq Profiling of human testes reveals the presence of the ACE2 receptor, a target for SARS-CoV-2 infection in spermatogonia, leydig and sertoli cells. *Cells* 2020;9:920.
53. Chandler JC, Bevins SN, Ellis JW, et al. SARS-CoV-2 exposure in wild white-tailed deer (*Odocoileus virginianus*). 2021; *bioRxiv*. 2021.07.29.454326. <https://doi.org/10.1101/2021.07.29.454326>
54. Cool K, Gaudreault NN, Morozov I, et al. Infection and transmission of ancestral SARS-CoV-2 and its alpha variant in pregnant white-tailed deer. 2021; *bioRxiv*. 2021.08.15.456341. <https://doi.org/10.1101/2021.08.15.456341>
55. Zhang G, Cowled C, Shi Z, et al. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 2013;339:456–460.
56. Jebb D, Huang Z, Pippel M, et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature*. 2020;583:578–584.

57. Dong D, Lei M, Hua P, et al. The genomes of two bat species with long constant frequency echolocation calls. *Mol Biol Evol.* 2017;34:20–34.
58. Ziegler CGK, Allon SJ, Nyquist SK, et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* 2020;181:1016–1035.e19.
59. Ito T, Suzuki Y, Mitnaul L, Vines A, Kida H, Kawaoka Y. Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. *Virology* 1997;227:493–499.
60. Emms DM, Kelly S. Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:1–14.
61. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–3587.e29.
62. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov.* 2021;2:100141.
63. Zhang Z, Zhu Z, Chen W, et al. Cell membrane proteins with high N-glycosylation, high expression and multiple interaction partners are preferred by mammalian viruses as receptors. *Bioinformatics* 2019;35:723–728.
64. Travaglini KJ, Nabhan AN, Penland L, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020;587:619–625.
65. Wang K, Zhu L, Sun Y, et al. Structures of Echovirus 30 in complex with its receptors inform a rational prediction for enterovirus receptor usage. *Nat Commun.* 2020;11.
66. Xiaofei E, Meraner P, Lu P, et al. OR14I1 is a receptor for the human cytomegalovirus pentameric complex and defines viral epithelial cell tropism. *Proc Natl Acad Sci USA.* 2019;116:7043–7052.
67. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015;16.
68. Hu H, Miao Y-R, Jia L-H, et al. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* 2019;47:D33–D38.
69. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 2010;5.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chen D, Ou Z, Zhu J, et al. Screening of cell-virus, cell-cell, gene-gene crosstalk among animal kingdom at single cell resolution. *Clin Transl Med.* 2022;12:e886.
<https://doi.org/10.1002/ctm2.886>