Molecular Biology and Evolution, 2026, 43, 1–16
https://doi.org/10.1093/molbev/msag072
Published: 5 May 2026
Discoveries

MBE OXFORD

Avian lung single-cell atlas elucidates evolutionary divergence in endothermic respiration

Yan Hao [1†], Xiang Zhou [2†], Qingshuo Zhao [1†], Huishang She [1], Yun Fang [1], Laikun Ma [3], Zhengting Zou [1], Weiwei Zhai [1,4], Per G.P. Ericson [5], Fumin Lei [1,4]*, Fanwei Meng [1]*, Shihua Zhang [2,6,7]*, Yanhua Qu [1,4,5]*

[1]State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[2]State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[3]Chengde Innovation Center for Exploitation and Ecological Protection of Distinctive Biological Resources, School of Biology and Food Science, Hebei Minzu Normal University, Chengde 067000, China
[4]College of Life Science, University of Chinese Academy of Sciences, Beijing 100049, China
[5]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, PO Box 50007, Stockholm SE-104 05, Sweden
[6]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[7]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[†]These authors contributed equally.
***Corresponding authors:** E-mails: quyh@ioz.ac.cn; zsh@amss.ac.cn; mengfw@ioz.ac.cn; leifm@ioz.ac.cn.
Associate editor: Guang Yang

Abstract

The evolution of distinct lung architectures in birds (tubular air capillaries) and mammals (alveoli) represents a classic example of convergent evolution, yet their cellular differences remain poorly characterized. We present the comparative single-cell atlas of avian and mammalian lungs, revealing four key findings. First, we discover a persistent hybrid cell population (AT1/AT2) in neonatal and adult birds that disappears postnatally in mice, uncovering distinct strategies for lung maintenance. Second, cross-species comparisons highlight striking divergence in alveolar cell types, reflecting specialized adaptations to different respiratory demands. Third, avian lungs show unique molecular signatures in proliferation and saccular development pathways, potentially explaining their superior respiratory efficiency. Fourth, our viral receptor analysis uncovered differential *ACE2* (SARS-CoV-2 receptor) expression but conserved *EGFR* (influenza A receptor), suggesting species-specific disease vulnerabilities. This work provides unprecedented insights into the evolutionary and developmental mechanisms shaping lung diversity in endotherms, with implications for respiratory biology, regenerative medicine, and zoonotic disease research.

Keywords single-cell RNA-sequencing, avian lung, cross-species comparison

Introduction

The evolution of specialized respiratory systems has been a critical adaptation enabling endotherms (i.e. birds and mammals) to sustain high metabolic rates and occupy a wide range of terrestrial environments (Hillenius and Ruben 2004). Although both groups share the physiological imperative for efficient gas exchange, they have evolved fundamentally divergent pulmonary structures. Birds have developed a complex system of rigid, tubular air capillaries that support unidirectional airflow and cross-current gas exchange. This arrangement permits more thorough oxygenation of the blood and is exceptionally effective at meeting the high aerobic demands of flight (Powell and Hopkins 2004). By contrast, mammals possess alveolar lungs that operate via tidal ventilation, in which air flows in and out along the same bronchial pathways. This bidirectional movement results in the mixing of fresh and residual air, which limits gas exchange efficiency. Consequently, the independent evolution of avian and mammalian lungs presents a pronounced functional convergence toward high-performance respiration accomplished through distinctly different structural and aerodynamic mechanisms.

The performance of the avian respiratory system arises from a set of highly specialized adaptations. Its flow-through architecture, comprising rigid parabronchial tubes and air capillaries, enables continuous, unidirectional gas exchange. This function is made possible by a clear anatomical division: Ventilation is powered by air sacs, whereas gas exchange occurs within the rigid, noncompliant lungs. When combined with cross-current blood

Received: June 9, 2025. **Revised:** February 3, 2026. **Accepted:** March 13, 2026

flow, this design yields high oxygen extraction efficiency, which is essential for supporting the metabolic requirements of flight (Monge and Leon-Velarde 1991; Ivy and Scott 2015). In mammals, however, respiration depends on bidirectional tidal ventilation, in which air follows the same route during both inhalation and exhalation, leading to the mixing of fresh and deoxygenated air. Additionally, the mammalian lung employs elastic, sac-like alveoli that serve dual roles in ventilation and gas exchange. Here, increased respiratory surface area is achieved mainly through alveolar multiplication rather than through structural separation of function (Powell and Hopkins 2004).

A comprehensive understanding of endothermal lung evolution requires not only cross-species comparison between birds and mammals but also an examination across developmental stages (Carroll 2005). The transition from neonatal to adult stages involves profound structural and functional maturation that is critical to an organism's respiratory capacity. In the neonatal lung, the architecture is often still developing; in mammals, this involves the formation and septation of alveoli to expand the gas exchange surface (Mund et al. 2008; Schittny 2017), while in birds, the intricate network of air capillaries undergoes extensive remodeling and vascularization to achieve its mature, rigid flow-through structure (Maina 2003; Makanya and Djonov 2009). Consequently, molecular programs, active in neonates, governing cell proliferation, differentiation, and tissue morphogenesis, differ significantly from those maintaining homeostasis and facilitating repair in adults (Morrisey and Hogan 2010; Kotton and Morrisey 2014).

The lung's function in gas exchange also makes it a critical interface for zoonotic transmission of respiratory viruses like influenza and coronaviruses (Clementi et al. 2021). The risk of cross-species transmission is influenced by a combination of species-specific cellular architecture and the distribution of viral receptors, a mechanism demonstrated by influenza virus recombination in swine that facilitates the emergence of novel strains like H5N1 and H7N9 (Rumschlag-Booms and Rong 2013; Harrington et al. 2021). Viral tropism is not uniform but depends on lung cellular heterogeneity, where receptors exhibit distinct patterns, from narrow cell-type specificity to broad, species-variable expression (Chen et al. 2021). Consequently, a single-cell resolution is indispensable for pinpointing vulnerable cell populations, identifying potential animal reservoirs, and accurately predicting zoonotic risks. This approach is vital for deciphering how the fundamental structural contrasts between avian and mammalian lungs create unique host–pathogen interaction landscapes.

Despite clear anatomical and physiological distinctions, the underlying cellular and molecular mechanisms directing the development and function of these systems remain poorly resolved. This knowledge gap is exacerbated by research focusing on domestic species and adult stages, leaving the developmental molecular biology of wild birds largely unexplored (Chen et al. 2021). Therefore, a comparative analysis of both neonatal and adult stages in a wild bird model is essential to delineate evolutionarily conserved pathways from lineage-specific adaptations, thereby providing a foundational cellular and molecular framework for understanding respiratory evolution in endotherms.

This study employs single-cell RNA sequencing (scRNA-seq) to construct the first comprehensive cellular atlas of the lung in a wild bird, the Eurasian tree sparrow (*Passer montanus*), spanning neonatal to adult developmental stages. We perform a cross-species comparative analysis to delineate the divergent cellular architectures and molecular programs underlying the evolution of the avian and mammalian respiratory systems. A key focus is the characterization of viral receptor distribution—including the SARS-CoV-2 receptor *ACE2* and the influenza receptor *EGFR*—revealing species-specific expression patterns that inform differential susceptibility. Our findings provide novel, cell-level insights into the evolutionary diversification of lungs in endotherms and the molecular determinants of zoonotic risk.

Results

Single-cell atlases for neonatal and adult tree sparrows reveal persistent hybrid cell type in avian lungs

To construct single-cell transcriptomic atlases of avian lungs, we generated scRNA-seq data from lung tissues of neonatal ($n = 6$) and adult ($n = 7$) tree sparrows (Fig. 1a; Tables S1 and S2). Following stringent quality control (see Materials and methods), we obtained 280,344 lung cells (Fig. S1 and Table S3). Utilizing cell marker genes for avian and mammalian lungs (i.e. Raredon et al. 2019; Chen et al. 2021; Table S4), we identified 17 cell types (Fig. 1b and c). Unsupervised clustering organized these cell types into four major clusters (Fig. 1c). The first cluster included alveolar type I cells (AT1s), type I/II cells (AT1/AT2s), type II cells (AT2s), pulmonary neuroendocrine cells (NECs), and ciliated cells (CCs); the second cluster comprised fibroblasts (FBs), pericytes (PCs), smooth muscle cells (SMCs) and endothelial cells (ECs); the third cluster consisted of six immune cell types, i.e. monocytes (MNCs), macrophages (MPs), dendritic cells (DCs), mast cells (MSCs), T cells (TCs), and B cells (BCs); and the fourth cluster included thrombocytes (THCs) and red blood cells (RBCs) (Fig. 1b and c). Cluster integrity was maintained across individuals (Figs. S3 and S4).

Among the three alveolar cell types, AT1 cells expressed *LIPM* and *SEMA3C*; AT2 cells expressed *FGL1* and *SFTPC*. Besides these two generally recognized alveolar cell types, we also identified a distinct cluster of cells co-expressed both AT1 and AT2 markers, which we termed "AT1/AT2s" (Fig. 1b and c; Fig. S2). AT1/AT2s not only co-expressed both AT1 and AT2 markers, also selectively expressed their own signature genes, including *PLCXD1* and *NDRG4*. Our Gene Ontology (GO) enrichment analysis of the cell-type-specific expressed genes for these alveolar cell types confirmed their unique biological functions. For example, GO term such as "morphogenesis of an epithelium" was enriched for AT1s; "cholesterol biosynthetic process" for AT2s, and "development growth" and "regulation of cell migration" for AT1/AT2s (Fig. S5).

We subsequently implemented pseudotime trajectory analysis (Cao et al. 2019) using Monocle3 to illustrate trajectory of the hybrid AT1/AT2 cell type in relation to AT1 and AT2 cell types (Fig. S6a). The results exhibited that the canonical markers for AT1 (*SEMA3C*) and AT2 (*SFTPC*) cells were co-expressed in AT1/AT2 hybrid cells during early pseudotime, and subsequently became highly enriched in their respective cell types by the middle and late pseudotime (Fig. S6b and c). These findings reveal a developmentally persistent AT1/AT2 hybrid cell population in the avian lung, as compared to AT1/AT2 hybrid cell type has only

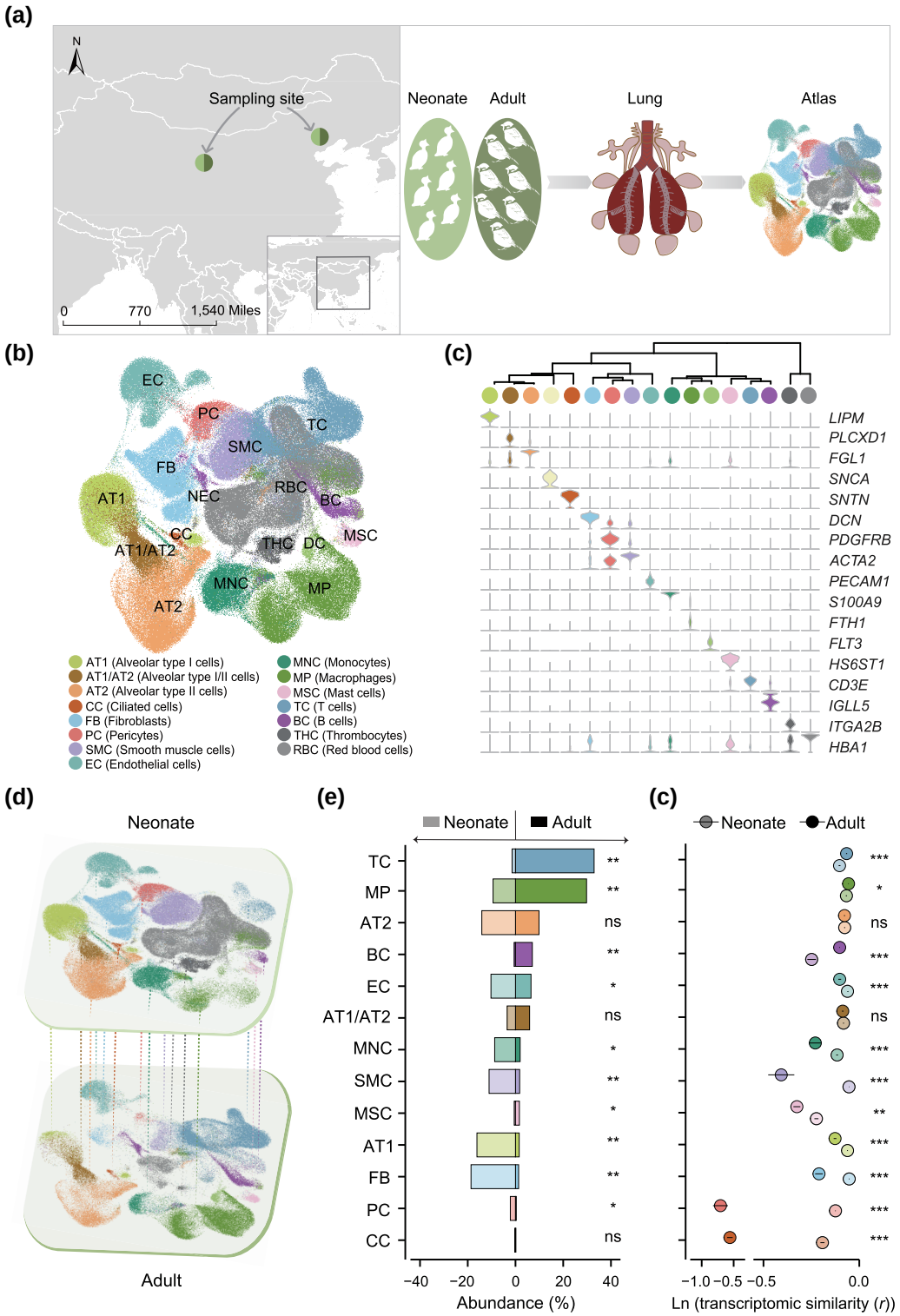

Figure 1 Construction of lung single-cell atlases and transcriptomic divergence between neonatal and adult tree sparrows. a) A schematic plot showing the lung single-cell atlases characterized for tree sparrows. The neonatal and adult tree sparrows were collected from two localities. b) The single-cell lung atlas identified 17 cell types, which were clustered into four major groups. AT1, alveolar type I cells; AT1/AT2, alveolar type I/II cells; AT2, alveolar type II cells; NEC, neuroendocrine cells; CC, ciliated cells; FB, fibroblasts; PC, pericytes; SMC, smooth muscle cells; EC, endothelial cells; MNC, monocytes; MP, macrophages; DC, dendritic cells; MSC, mast cells; TC, T cells; BC, B cells; THC, thrombocytes; RBC, red blood cells. c) Cell marker genes exhibited high enrichment scores for the lung cell types. d) The lung cell types were well-aligned between neonatal and adult birds. e) Cell abundance of the lung cell types exhibited dramatic changes between neonatal and adult tree sparrows. f) Transcriptomic similarity based upon expression correlation of homologous lung cell types observed between neonatal and adult birds. Data were presented as median ± SEM. In e) and f), statistical significance was tested using Wilcoxon rank-sum test with false discovery rate correction for multiple comparisons: ns, non-significant; *, adjusted $P < 0.05$; **, adjusted $P < 0.01$; ***, adjusted $P < 0.001$.

previously been reported in embryonic (E16.5) and neonatal (P1) mice (Desai et al. 2014; Guo et al. 2019). Taken together, these results, to our best knowledge, establish the first single-cell atlases of lung tissues across developmental stages in a wild bird species, further expanding previously generated lung single-cell data from domestic birds such as ducks and pigeons (Chen et al. 2021).

Transcriptional heterogeneity of avian lung across developmental stages

Our reconstructed single-cell atlases revealed conservation of lung cell types between neonatal and adult tree sparrows. These homologous cell types from both developmental stages showed close clustering in Uniform Manifold Approximation and Projection (UMAP) layouts (Fig. 1d; Fig. S7) with consistent marker gene expression patterns (Fig. S2). Comparative analysis identified significant changes in cell-type proportions between developmental stages. Neonatal lung exhibited higher fractions of structural components including AT1 cells, FBs, PCs, and SMCs. In contrast, adult lungs showed expanded immune populations, particularly MPs, MSCs, TCs, and BCs (Fig. 1e; Table S5). These compositional changes likely reflect the structural maturation of lung development and the adult's increased need for immune surveillance.

Gene expression analysis revealed distinct regulatory patterns between neonatal and adult avian lungs. Most cell types displayed greater transcriptional homogeneity in neonatal lungs (Fig. 1f; Table S6), suggesting tighter control of developmental programs. Among immune cells, B cells showed the inverse pattern, with higher expression consistency in adults, potentially indicating functional specialization of the mature immune system. These findings collectively demonstrate how avian lungs maintain core cellular architecture while dynamically adjusting their composition and gene expression programs. In addition, the observed changes from structural growth in neonates to adaptive immune competence in adults correspond to shifting physiological demands.

Cellular heterogeneity and transcriptomic divergence of lungs across species and developmental stages

Despite the distinct morphological organization of avian lungs compared to mammalian lungs (Fig. 2a), our cross-species analysis revealed striking conservation of cellular composition. Using CAME (Liu et al. 2023) to map cell types between tree sparrows, mice, and humans, we found high matching rates for both neonates (84.3% to 98.8% for tree sparrow/human, 92.2% to 99.1% for tree sparrow/mouse) and adults (69.6% to 98.7% for tree sparrow/human, 64.3% to 99.9% for tree sparrow/mouse; Fig. 2b). These mapping rates between tree sparrows and either mouse or human were similar to those between the mouse and human lungs, as exhibited by 70.2% to 99.9% matching rates for adults and 79.0% to 99.4% for neonates (Fig. S8). This suggests that core lung cell types are evolutionarily conserved across endotherms, despite their divergent respiratory structures.

To assess transcriptomic variation, we quantified cellular heterogeneity using entropy, which measures gene expression variability within cell populations (Ma et al. 2022). While AT1 and AT2 cells showed consistent heterogeneity levels across species and developmental stages, structural cell types, ECs, FBs, and SMCs, exhibited greater transcriptional diversity in neonates than adults (Fig. 2c). This is likely to reflect developmental plasticity in early life, with cell states becoming more constrained in mature lungs.

Comparative gene expression analysis also revealed cell-type-specific patterns of evolutionary divergence. AT1 and AT2 cells showed the greatest transcriptomic differences (1-Spearman's coefficient) between birds and mice or humans (Fig. 2d). In contrast, FBs and ECs were more conserved, suggesting that core stromal functions are under strong evolutionary constraints. For all these cell types, the transcriptomic differences between mouse and human lungs were smaller than between sparrows and either mouse or human lung, with the difference across species pairs being more significant in FBs (e.g. 0.32 and 0.44 for adult and neonatal mouse/human vs. 0.52 and 0.52 for both adult and neonatal human/sparrow and 0.45 and 0.49 for adult and neonatal mouse/sparrow; Table S7). These trends were consistent across neonates and adults, indicating that species-specific adaptation arises early in development. Collectively, these results suggest that while lung cell types are largely homologous across endotherms, their transcriptional programs have diverged to different extents, with a great extent in AT cells.

Cross-species conservation of lung cell-type gene expression

Our analysis of 4,887 to 5,675 orthologous genes revealed fundamental evolutionary patterns in lung cell-type gene expression across tree sparrows, mice, and humans. Hierarchical clustering of major cell types (e.g. AT1s, AT2s, FBs, ECs, and CCs) consistently grouped samples according to phylogenetic relationships, with mice and humans forming one cluster distinct from tree sparrows (Fig. S9). This clear separation was observed consistently across different developmental stages and was further supported by principal component analysis (PCA; Fig. S9), demonstrating that avian and mammalian lung cells maintain distinct transcriptional profiles.

Using *k*-means clustering and elbow methods, we identified sets of highly expressed, evolutionarily conserved genes for each major lung cell type: 613 genes for AT1 cells, 559 for AT2 cells, 487 for CCs, 567 for FBs, and 581 for ECs, respectively (Fig. 3a; Figs. S10 to S14). Approximately 50% of these conserved genes were shared across different cell types (Fig. S15), with significant functional enrichment for essential metabolic processes including aerobic respiration, oxidative phosphorylation, and mitochondrial ATP synthesis (Fig. 3b). Two particularly notable classes of conserved genes emerged from our analysis. First, multiple aerobic respiration-related genes (e.g. *CHCHD2*, *NDUFA4*, *COX6C*, and *ATP5F1B*) showed consistently high expression across all cell types and species.

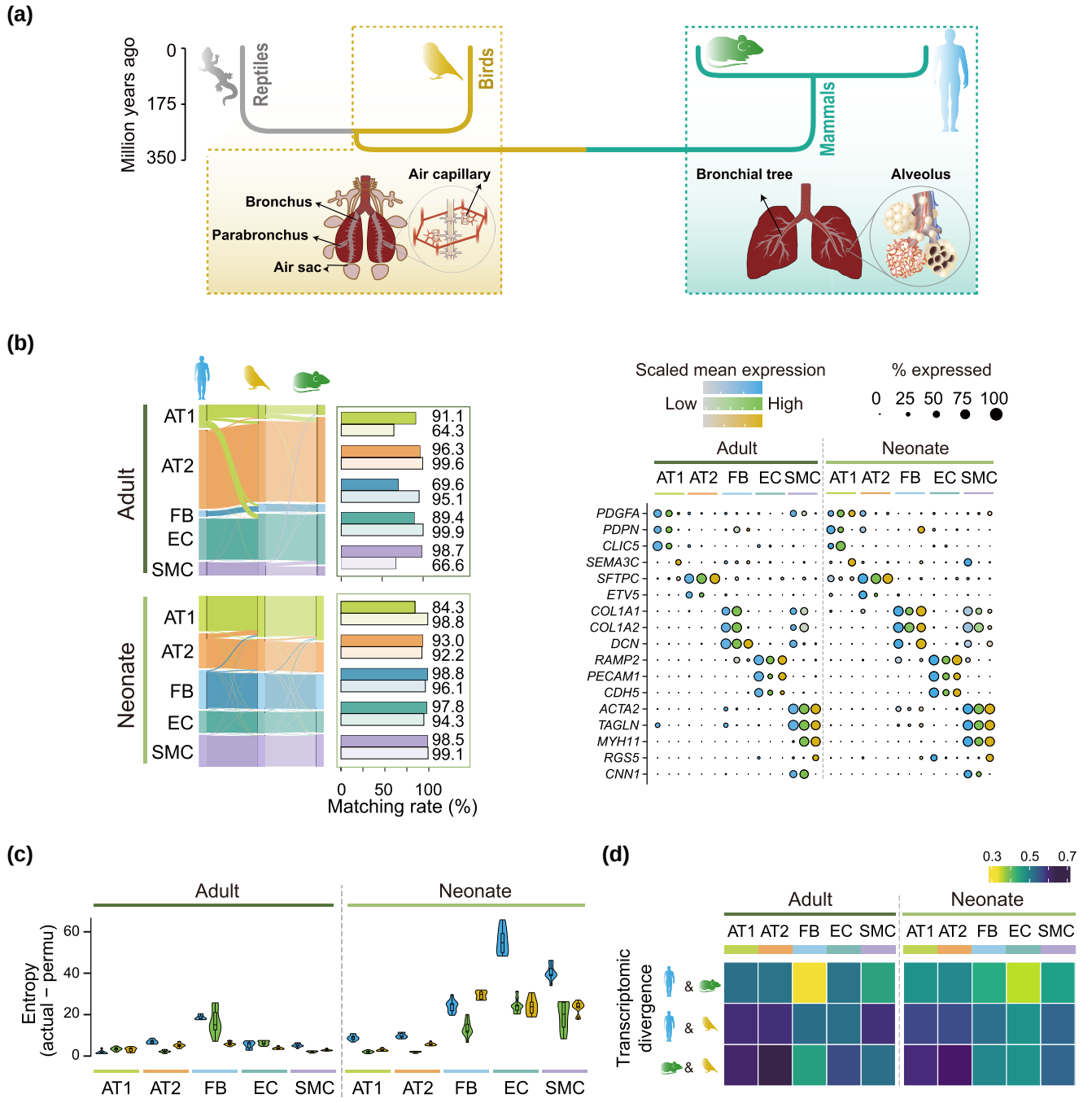

Figure 2 Cell-type heterogeneity and transcriptomic divergence of lung tissues across birds, mice, and humans. a) Birds and mammals evolved divergent lungs, exemplified by the air sac lungs of birds compared to the alveolar lungs of mammals. Reptile (gray) in the phylogenetic tree was only used to illustrate the parallel evolution of avian and mammalian lungs after the avian–reptilian and the mammalian lineages split, but not included in this study. b) The lung cell types exhibited high alignment across tree sparrows, mice, and humans. The matching rates of cell types between tree sparrows and mice or humans (left) and the expression of cell-type marker genes across the species (right) were depicted. c) Cellular heterogeneity of each lung cell type was indicated by entropy, measuring the overall transcriptomic heterogeneity among cells within a cell type. d) Transcriptomic divergence in the homologous lung cell types was assessed. In b) to d), neonates and adults were presented separately.

Second, several calcium-dependent phospholipid-binding proteins (*ANXA2* and *ANXA5*) demonstrated similar patterns of conserved expression (Fig. 3c and d). The strong evolutionary conservation of these genes reflects their critical roles in maintaining basic cellular functions including energy metabolism, cellular growth regulation, and signal transduction pathways (Spinelli and Haigis 2018; Gerke et al. 2024).

Cross-species comparisons of lung cell types exhibited avian-enriched gene expression

Given the structural differences between avian tubular air-capillary lungs and mammalian alveolar lungs, we identified distinct, species-specific gene expression patterns in tree sparrows.

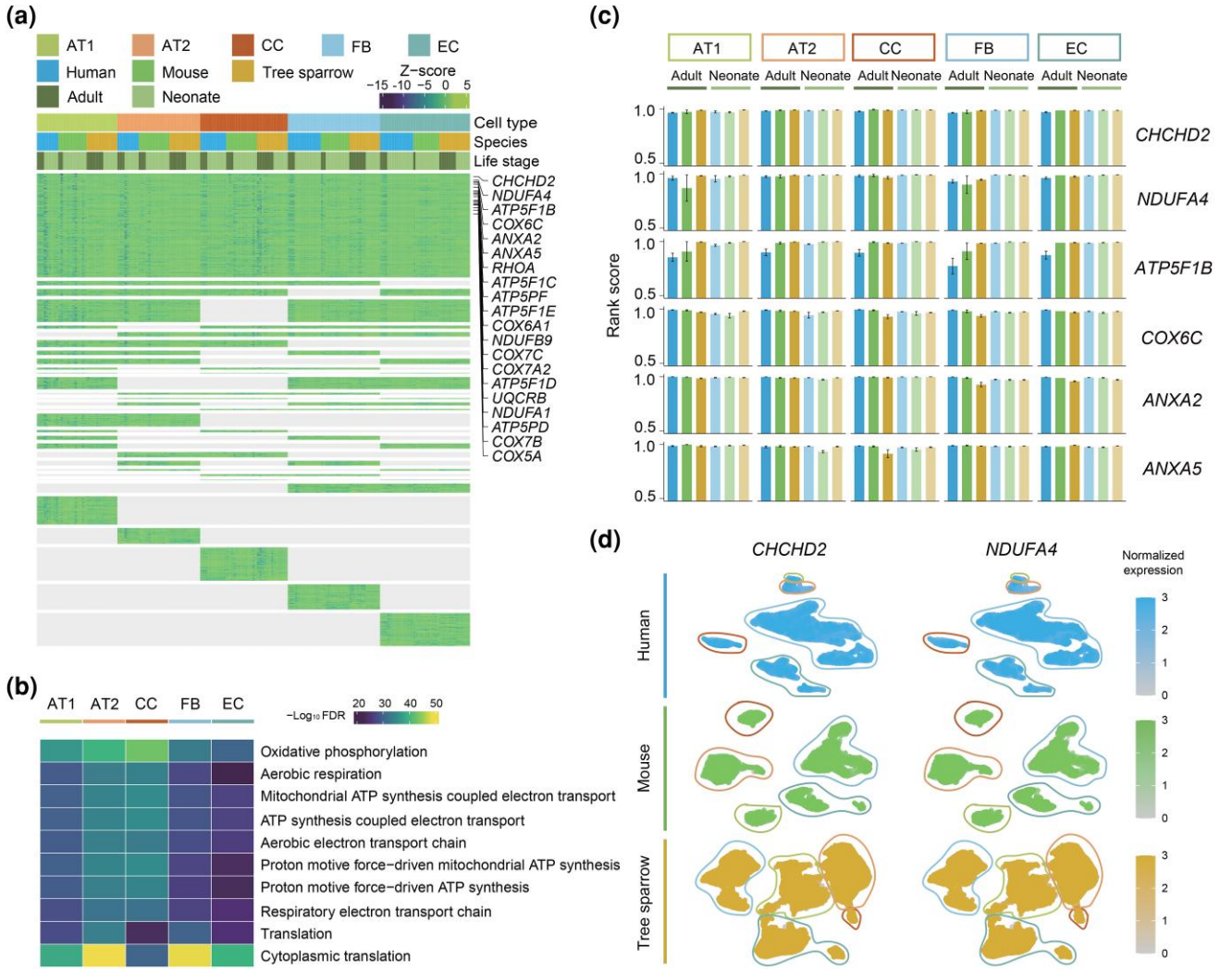

Figure 3 Conserved gene expression features of lung cell types across species. a) Conserved gene expression programs of the lung tissue among tree sparrows, mice, and humans. b) Conserved genes were enriched for aerobic respiration, oxidative phosphorylation, and mitochondrial ATP synthesis. c) The exemplary genes that were conservedly expressed across lung cell types and different species. d) *CHCHD2* and *NDUFA4* were uniformly expressed in all lung cell types of tree sparrows, mice and humans.

Focusing on avian-enriched transcriptional programs, we detected 479 differentially expressed genes in AT1 cells and 748 genes in AT2 cells (Figs. S10 and S11), the two cell types exhibiting the greatest transcriptomic divergence across the species studied (Fig. 2d). Functional enrichment analysis indicated that these genes are involved in cell migration, metabolic, and catalytic activities (Fig. S16).

Among the most highly expressed genes unique to tree sparrows, *PSCA* (encoding a glycosylphosphatidylinositol-anchored cell membrane glycoprotein) and *LYZ* (encoding an essential component of lung innate defense) showed broad enrichment across AT1, AT2, CC, FB, and EC cells (Fig. 4a and b). Notably, *PSCA*, which influences cell proliferation and serves as a diagnostic marker in human lung cancer (Kawaguchi et al. 2008; Wei et al. 2017), was selected for further functional investigation in an avian cellular context.

To address this, we overexpressed tree sparrow *PSCA* (accession XM_039714568.1) in chicken DF-1 fibroblasts. Cell-cycle analysis via flow cytometry (BD Fortessa) comparing *PSCA*-expressing cells with empty-vector controls revealed that *PSCA* overexpression significantly inhibited proliferation. Quantification of DNA content using ModFit software showed a pronounced decrease in the proportion of cells in S and G2/M phases, alongside a concomitant increase in G0/G1 populations (Fig. 4c). Consistent with this result, EdU fluorescence staining confirmed a significant suppression of proliferation in *PSCA*-expressing cells relative to controls (Fig. 4d).

We further investigated the cell–cell interactions enriched to tree sparrows, identifying 14 tree sparrow-enriched signaling pathways that were conserved between adult and neonatal lungs (Fig. 4e; Figs. S17 and S18). We also detected 12 adult-enriched and 47 neonatal-enriched interaction events in tree sparrow lung (Figs. S19 and S20). These pathways include lipid transport (*APOA1* to *ABCA1*), guidance signaling (SEMA3A-NRP1 and SEMA3A-PlexinA1), growth regulation (TGF-β, *TGFB2* to *TGFBR2*; FGFR, *CD44* to *FGFR2*), extracellular matrix interactions (laminin, *FAM3C* to *LAMP1*), integrin-vitronectin binding (*VTN* to *ITGAαVβ1*), and collagen-binding integrin pathways (*COL4A1*/*COL4A2* to *ITGAα1β1*). To validate these predictions in situ, we selected the top-ranked avian-enriched pair

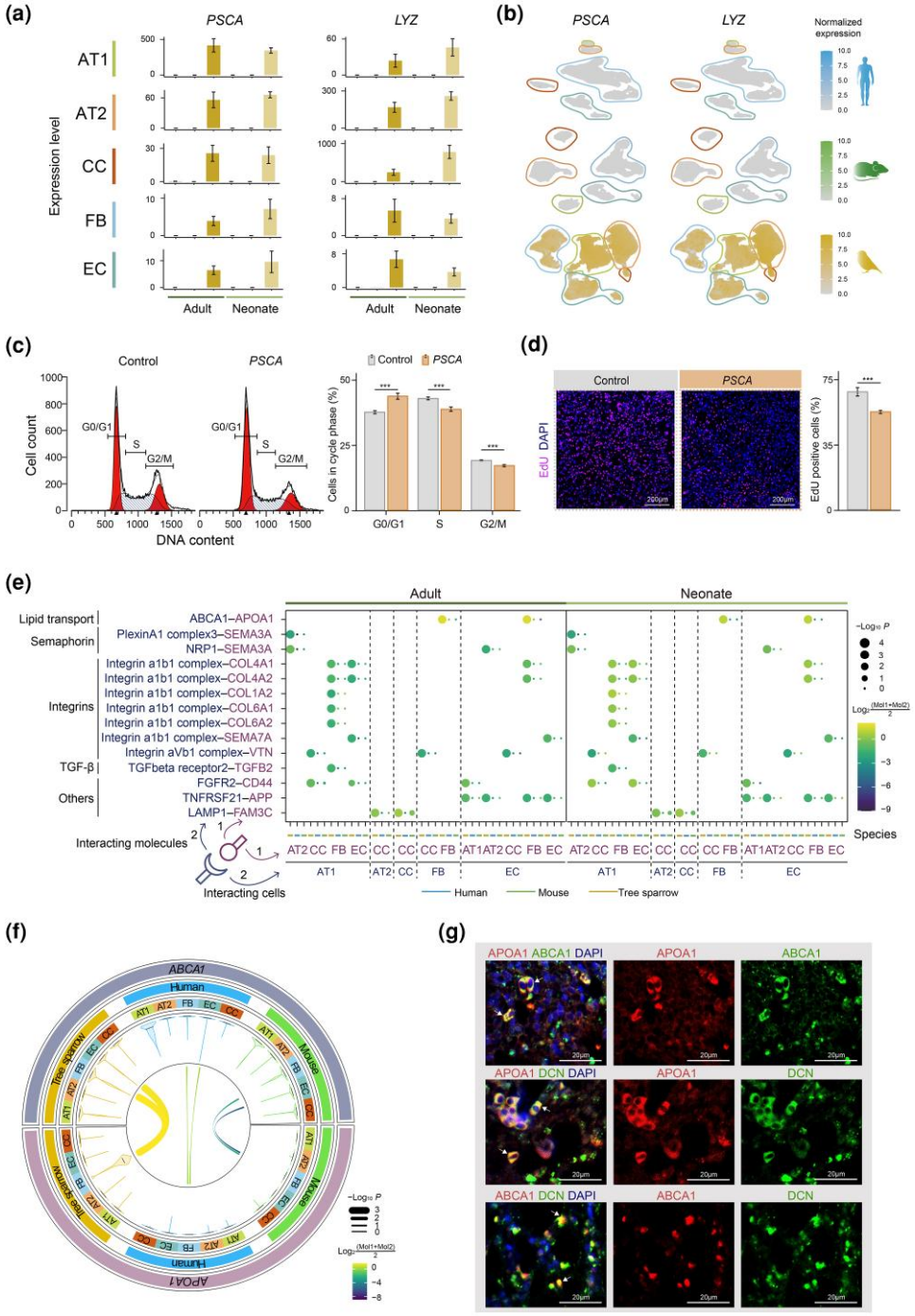

Figure 4 Avian-enriched gene expression and cell–cell interaction. a) *PSCA* and *LYZ* were highly expressed in lung cell types of tree sparrows but not in the lungs of mice and humans. b) *PSCA* and *LYZ* were widely expressed in all lung cell types of tree sparrows but not in the lungs of mice and humans. c) Flow cytometric analysis of cell-cycle distribution in chicken DF-1 fibroblasts based on DNA content following propidium iodide (PI) staining. Left panels show representative DNA content histograms with model-fitted G0/G1, S, and G2/M phases in control and *PSCA*-overexpressing cells. The right panel shows the percentage of cells in each cell-cycle phase. Quantitative analysis shows a significant increase in the proportion of cells in the G0/G1 phase and a concomitant decrease in the proportion of cells in S and G2/M phases upon *PSCA* overexpression compared with the control cells (data are presented as mean ± SEM, $n = 9$; adjusted $P < 0.001$). d) EdU incorporation assay to assess DNA synthesis in DF-1 cells. Representative immunofluorescence images show EdU-positive nuclei (magenta). Note that the Alexa Fluor 488 EdU signal was pseudo-colored red and merged with DAPI-stained nuclei (blue). Scale bars, 200 μm. Quantitative analysis of the EdU-positive rate confirms a significant reduction in DNA synthesis in *PSCA*-overexpressing cells compared with controls (mean ± SEM, $n = 13$; $P < 0.001$). e) Cell–cell interactions were observed in lung cell types of adult and neonatal tree sparrows but absent in mice and humans. f) Avian-enriched ligand–receptor pair *APOA1–ABCA1* in tree sparrow lung. g) Representative dual-color FISH images of adult tree sparrow lung sections. Rows display co-staining combinations of *APOA1* (red) and *ABCA1* (green) (top), *APOA1* (red) and *DCN* (green) (middle), and *ABCA1* (red) and *DCN* (green) (bottom). Left panels show merged images with DAPI nuclear counterstain (blue); middle and right panels show individual fluorescence channels. White arrows indicate co-localization of the two signals (yellow). Scale bars, 20 μm.

APOA1-ABCA1 and performed dual-color fluorescent in situ hybridization (FISH) on adult tree sparrow lung sections (Fig. 4f). This analysis confirmed co-localization of *APOA1* and *ABCA1* transcripts within *DCN*-positive fibroblasts (Fig. 4g), a spatial expression pattern consistent with CellPhoneDB prediction and supportive of a local lipid-transport process in avian lung. Collectively, these results identify an enrichment of gene regulatory programs and communication networks that are distinctive to birds, which could provide a plausible mechanistic foundation for the evolution of the specialized avian pulmonary structure.

Cell-type-specific expression of respiratory virus receptor genes in avian and mammalian lungs

The lung serves as a primary target tissue for respiratory viruses, with its cellular diversity playing a crucial role in determining host susceptibility (Clementi et al. 2021). To investigate how the distinct organizations of avian and mammalian lungs affect viral receptor distribution, we analyzed the expression of 18 viral receptor genes (Ramos and Fernandez-Sesma 2012; Pirooz et al. 2014) across tree sparrows, mice, and humans (Fig. 5a). Our analysis revealed two distinct expression patterns that provide important insights into species-specific viral susceptibility. Several key viral receptors showed broad expression across all major lung cell types (i.e. AT1s, AT2s, CCs, FBs, and ECs) in all three species examined. These included *ITGAV*, which serves as the entry receptor for adenovirus type C (human parechovirus 1), and *ITGB1*, the receptor for mammalian reovirus (human parvovirus B19). Particularly noteworthy was *EGFR*, the influenza A virus receptor, which was consistently expressed in all major cell types across birds and mammals (Fig. 5b). Immunofluorescence staining confirmed *EGFR* expression in the lungs of both tree sparrows and mice (Fig. 5c).

In contrast to these broadly expressed receptor genes, we observed striking species-specific differences in other viral receptors. Most significantly, angiotensin-converting enzyme 2 (*ACE2*), the established receptor for SARS-CoV-2, showed markedly different expression patterns. While *ACE2* was readily detected in AT2 and ciliated cells of human and mouse lungs, it was absent from all lung cell types in tree sparrows (Fig. 5d). Immunofluorescence validation confirmed abundant *ACE2* expression in mouse lungs but failed to detect any signal in lung tissue of tree sparrows (Fig. 5e and f). Together, our results highlighted that distributions of respiratory virus receptor genes exhibit species specificity.

Discussion

Understanding the evolution of avian and mammalian lungs requires a comparative analysis of their cellular and molecular compositions. Here, we present the first single-cell transcriptomic atlas of the lung spanning neonatal and adult stages in a wild bird. This work expands upon previous studies, which have been confined to domestic species such as ducks and pigeons (Chen et al. 2021). Through a comparative analysis of our avian data with existing datasets from mice and humans, we reveal key evolutionary distinctions. First, we identify the retention of an AT1/AT2 hybrid cell type into adulthood in birds, unlike in mice, where this cell population disappears after birth. Second, we observed that the extent of transcriptional divergence between avian and mammalian lungs varied by cell types, with AT1 and AT2 exhibiting the greatest divergence. Third, by integrating comparative single-cell analysis and immunofluorescence validation, we demonstrate that avian lungs lack *ACE2* expression (linked to SARS-CoV-2 susceptibility) while maintain *EGFR* (linked to influenza susceptibility). Collectively, these findings provide novel insights into evolutionary trajectories of avian and mammalian lungs.

Our comparative analysis of lung cell types across birds, mice, and humans identified a persistent AT1/AT2 hybrid cell type in tree sparrows, present at both neonatal and adult stages. This finding stands in contrast to mice, where this cell type has been detected only during embryonic and early postnatal development (Desai et al. 2014; Guo et al. 2019). In mice, the hybrid cell type is considered a bipotent progenitor; however, its functional potential in tree sparrows remained unknown. Our pseudotime trajectory analysis indicates that the avian AT1/AT2 hybrid cells can differentiate into mature AT1 or AT2 cells (Fig. S6), suggesting it retains bipotent progenitor potential. To conclusively validate its multi-lineage differentiation capacity and regenerative function, future work should employ functional studies, such as in vitro differentiation assays and lineage tracing. Ultimately, a direct comparison of the regulation and function of this persistent hybrid cell in birds versus mice could yield novel insights into the evolutionary mechanisms underlying lung repair and regeneration.

We observed that the extent of transcriptional divergence varied among cell types across species, with alveolar type cells exhibiting the greatest divergence. This substantial molecular divergence likely reflects fundamental differences in lung structure and physiology. AT1 cells form the gas exchange surface, and in birds, this occurs within rigid air capillaries with a thinner, more uniform blood-gas barrier than the compliant alveoli of mammals (Maina 2007; West 2009; Desai et al. 2014). Similarly, AT2 cells produce pulmonary surfactant; however, its composition and function are distinct. Avian surfactant is uniquely adapted to maintain air-capillary patency, contrasting with the mammalian surfactant's primary role in preventing alveolar collapse (Bernhard et al. 2001; Whitsett et al. 2010). Taken together, these pronounced transcriptional differences in AT1 and AT2 cells may underscore their specialized roles in the distinct respiratory architectures of birds and mammals. Future work should characterize the specific gene expression patterns of these cell types to determine their contribution to the distinct architectures of the avian and mammalian lung.

The identification of avian-specific molecular signatures advances our understanding of the genetic underpinnings of lung function divergence. We found that *PSCA* is broadly and highly expressed in multiple pulmonary cell types in tree sparrows. Interestingly, in humans, this gene is recognized as biomarkers of lung cancer; its knockdown suppresses tumorigenesis (Kawaguchi et al. 2008; Wei et al. 2017). This presents a compelling paradox: How do avian lungs tolerate high expression of these potentially oncogenic molecules while maintaining tissue homeostasis? Our functional analysis in avian cell model revealed a critical divergence. Contrary to its established pro-

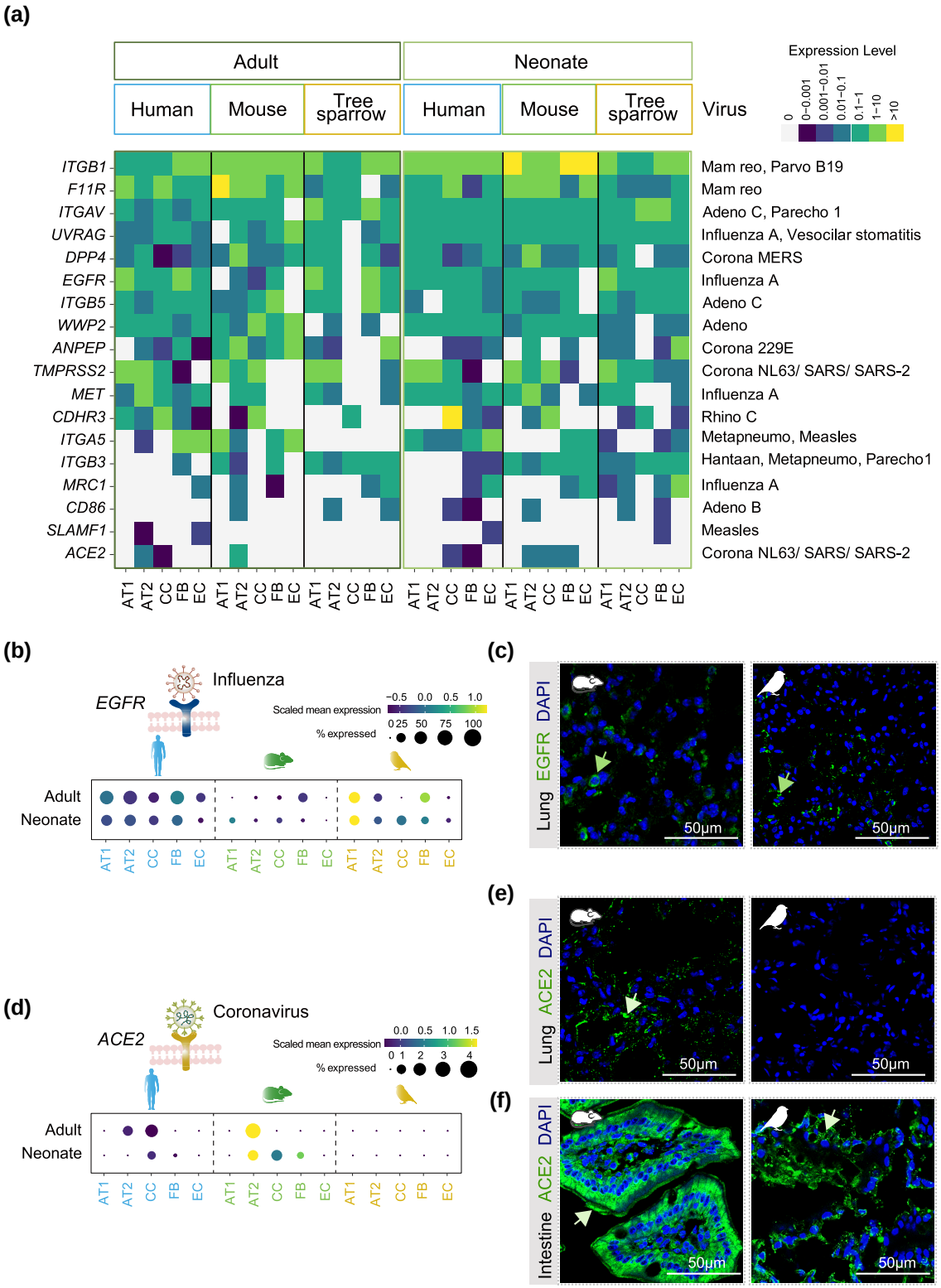

Figure 5 Expression patterns of respiratory virus receptors in lung cell types across species. a) Heatmap showing gene expression in lung cell types of entry receptors for respiratory viruses across tree sparrows, mice, and humans. b) Pan-expression of *EGFR* across species was observed. c) Immunostaining validation of EGFR in lungs from mice (left) and tree sparrows (right). Scale bars represent 50 μm. d) *ACE2* expression was absent in the lungs of tree sparrows but present in the lungs of mice and humans. e) Immunostaining validation of ACE2 in lungs from mice (left) and tree sparrows (right). f) Positive control for immunofluorescence staining of ACE2 in small intestines from mice (left) and tree sparrows (right). c), e), and f) Scale bars represent 50 μm.

proliferative role in human cancers, overexpression of avian *PSCA* inhibited cell proliferation and induced G0/G1-phase arrest, effectively blocking the G1/S transition. Collectively, these results provide a mechanistic insight into the aforementioned paradox: Avian lungs may tolerate high *PSCA* expression because, in this evolutionary context, it has been co-opted to promote cellular quiescence and tissue stability rather than unchecked proliferation. Future studies should investigate the specific regulatory mechanisms that enable this functional reversal in birds, which may reveal novel pathways for controlling cell proliferation with therapeutic potential for human diseases, for example, cancer.

Furthermore, we identified an enrichment of avian-specific molecular signatures and signaling pathways known to be critical for alveolar and saccular morphogenesis. For example, semaphorin signaling, a regulator of lung branching morphogenesis (Kagoshima et al. 2001), may be instrumental in patterning the unique tubular air-capillary network of the avian lung, which is structurally distinct from the mammalian alveoli (Bernhard et al. 2001). Similarly, TGF-β signaling, which governs essential epithelial-mesenchymal interactions during lung development (Saito et al. 2018), was also enriched. Notably, pathways involved in basement membrane organization were prominent. Collagen-binding integrins, such as ITGAα1β1, are vital for alveolar septation, as their deficiency causes defects in mice (Watanabe-Takano et al. 2024). The binding of integrin ITGAα1β1 to type IV collagen has been linked to cellular flattening (Tournier et al. 1992), and the expression of its ligands, *COL4A1* and *COL4A2*, is upregulated during the formation of the avian blood-gas barrier (Loscertales et al. 2016). This is highly relevant as birds possess exceptionally flattened AT1 cells that optimize gas exchange and support high aerobic capacity (Bjørnstad et al. 2014). While these pathways are functionally documented, the specific mechanisms by which their avian-specific orchestration contributes to the unique vascular networks of the tubular air-capillary lung remain unknown. Future studies are needed to clarify how these cell–cell interactions functionally underpin the remarkable respiratory efficiency of the avian lung.

Our profiling of viral receptor expression across species suggests distinct host–pathogen coevolution trajectories for different respiratory viruses. We found that conserved *EGFR* expression across avian and mammalian lungs correlates with universal susceptibility to influenza viruses, indicating that this pan-species receptor may facilitate zoonotic transmission. This is exemplified in pigs, where the coexistence of both avian-type (SAα2,3Gal) and human-type (SAα2,6Gal) receptors in the respiratory epithelium allows them to act as a "mixing vessel" for viral reassortment (Ito et al. 1998). Conversely, *ACE2* expression is absent in avian lungs but present in those of mice and humans. This finding aligns with the observed susceptibility to SARS-CoV-2, which infects mammals, including dogs, cats, and humans, but not birds such as chicken and ducks (Shi et al. 2020; Chen et al. 2021). Furthermore, the specialized structure of the avian lung likely creates an additional physical barrier by restricting viral deposition and accessibility to target cells (Maina 2007). Collectively, these findings underscore how interplay between specific receptor distribution and host respiratory structure shapes viral transmission. This opens a new avenue for research to experimentally test viral entry potentials across the divergent structures of avian and mammalian lungs.

Several limitations in the present study should be acknowledged. First, achieving perfect alignment of developmental stages across diverse species is challenging due to their distinct ontogenetic trajectories and life-history strategies (Cardoso-Moreira et al. 2019). Consequently, while we employed broad neonatal and adult categories for cross-species comparison, this approach cannot fully resolve finer-scale, stage-specific transcriptional differences. Second, although we tentatively infer that the AT1/AT2 hybrid cell population, which persists in both neonatal and adult birds, may possess progenitor potency, functional validation in a wild bird species was not feasible in this study. Future work using in vitro assays or lineage tracing is required to test the differentiation potential of this population in the avian lung.

Despite these limitations, our findings substantially advance the current understanding of avian lung evolution and provide a valuable genomic resource for future research. The avian lung could provide a powerful model for addressing fundamental questions in organ development, regeneration, and adaptation to physiological demands. Furthermore, its unique properties may offer novel perspectives in zoonotic disease ecology and may inform therapeutic strategies for human lung pathologies.

Materials and methods

Sample collection

We collected lung tissue samples from adult and neonatal (postnatal day 1) tree sparrows from two locations in China (Fig. 1a). Samples included three adults and three neonates from Chengde, Hebei, and four adults and three neonates from Gahai, Gansu (Table S1). All sample collections were conducted in full compliance with the Ethics Committee of the Institute of Zoology, Chinese Academy of Sciences. Our procedures strictly adhered to China's National Wildlife Conservation Law to ensure ethical treatment and minimal environmental impact.

Tissue processing and single-cell isolation

The lung tissues processing followed a standardized protocol to ensure high-quality single-cell suspensions. Initial preparation involved thorough washing of tissues with Hanks Balanced Salt Solution (HBSS) followed by careful dissection into 1 to 2 mm^3 fragments using sterile techniques. For cellular dissociation, we employed GEXSCOPETM Tissue Dissociation Solution (Singleron) with incubation at 37 °C for 15 min, accompanied by gentle mechanical dissociation through pipetting to achieve optimal cell separation.

Following enzymatic digestion, the cell suspension underwent sequential filtration through 40-μm sterile strainers to remove undigested tissue fragments. Centrifugation at 1,000 rpm (4 °C for 5 min) allowed supernatant removal and subsequent resuspension in 1× PBS. We implemented rigorous quality control measures, including viability assessment through trypan blue exclusion and microscopic examination of cell morphology. All samples were standardized to a minimum concentration of 1×10^5 cells per microliter (μL) to ensure sufficient cellular output for downstream single-cell RNA sequencing applications.

scRNA-seq

scRNA-seq library preparation was performed using the GEXSCOPE® Single Cell RNA Library Kit (Singleron) following the manufacturer's protocol (Dura et al. 2019). Briefly, single-cell suspensions were prepared at concentrations of $1\text{-}3 \times 10^5$ cells/μL in 1 × PBS and loaded onto microfluidic chips using the Singleron Matrix® Single Cell Processing System to partition individual cells into separate wells. Cell barcoding beads were then introduced into the chip and rinsed, followed by the addition of a lysis buffer to release and capture mRNAs at room temperature for 20 min. After retrieval of the barcoded beads and captured mRNAs, reverse transcription, cDNA amplification via PCR, and library construction were carried out. The final libraries underwent size selection, purification, and quality control before being sequenced on an Illumina NovaSeq 6000 platform with 150 bp paired-end reads. All scRNA-seq procedures were conducted by Singleron Biotechnologies (Nanjing, China).

Raw sequencing data processing

Sequencing data processing was performed using CeleScope v1.14.1 (https://github.com/singleron-RD/CeleScope). Briefly, reads lacking poly-T tails were filtered out, and low-quality sequences were trimmed using Cutadapt v3.7 (Martin 2011) to remove poly-A tails and adapter sequences. The cleaned reads were then aligned to the tree sparrow genome (https://db.cngb.org/search/project/CNP0006114/) using the splice-aware aligner STAR v2.6.1a (Dobin et al. 2013). Gene and UMI counts per cell were quantified with featureCounts v2.0.1 (Liao et al. 2014) to generate the final count matrices. Across all samples, the median number of genes per cell was 818 and the median UMI count per cell was 2,008, with an average sequencing saturation of 76% (Table S2). Detailed quality metrics and sequencing statistics were provided in Table S2.

Quality control

To ensure data quality, we performed contamination correction, doublet detection, and stringent cell filtering. First batch effects and ambient RNA contamination were addressed using SoupX v1.5.0 (Young and Behjati 2020). Genes exhibiting bimodal expression, enrichment in empty droplets, or association with specific cell groups were identified and functionally categorized. The contamination fraction (rho) was estimated based on these genes, and expression profiles were corrected accordingly. Next, doublets were detected using DoubletFinder v2.0.3 (McGinnis et al. 2019), where artificial doublets were generated by averaging randomly paired cell profiles (pN = 0.25). The optimal neighborhood size (pK) was determined using the mean-variance normalized bimodality coefficient (BCMVN), and doublets were predicted via the "doubletFinder_v3" function. A total of 19,692 doublets were identified and removed.

Further quality filtering was conducted using Seurat v4.0.1 (Hao et al. 2021). Cells were excluded if they met any of the following criteria: (i) fewer than 100 detected genes (nGenes <100), (ii) high mitochondrial gene content (mitoRatio ≥50%), (iii) excessive ribosomal gene expression (riboRatio ≥20%), and (iv) low transcriptional complexity (log10(GenesPerUMI) ≤ 0.7). Additionally, only genes detected in at least ten cells per sample were retained. Following quality control, the dataset retained 280,344 cells with a median of 744 genes and 1,822 UMIs per cell (Fig. S1 and Table S3). Post-QC metrics, including cell and gene retention rates, were summarized in Fig. S1 and Table S3.

Normalization, scaling, integration, clustering, and annotation

Data processing was performed using Seurat v4.0.1 (Ahlmann-Eltze and Huber 2021; Hao et al. 2021), following a standardized workflow. For normalization, UMI counts per gene were divided by the total UMI counts per cell, scaled by a factor of 10,000, and log-transformed using "NormalizeData". The "SCTransform" method was applied for variance stabilization and bias correction.

To account for biological variation, lung samples from neonates and adults were integrated. Highly variable genes (3,000) were selected using "SelectIntegrationFeatures", followed by anchor identification "PrepSCTIntegration" and "FindIntegrationAnchors". After integration, PCA was conducted using "RunPCA", and clusters were identified using "FindCluster", a shared near-neighbor (SNN) graph-based approach. For dimensionality reduction and visualization, uniform manifold approximation and projection (UMAP) was applied using "RunUMAP" with a correlation metric. Marker genes for each cluster were identified using "FindAllMarkers" with thresholds of adjusted $P < 0.05$ and $|\log2(\text{fold change})| > 1$. Cell types were annotated based on canonical marker genes (Table S4).

Functional enrichment analysis was performed on marker genes using ToppFun, querying the GO biological process database. Statistical significance was assessed via a hypergeometric test with Benjamini–Hochberg correction for multiple testing (Chen et al. 2009).

Trajectory inference of AT1/AT2 hybrid cell population

We isolated alveolar epithelial cells from the whole-cell dataset. Cells within each cell type were randomly down-sampled to equalize cell numbers across cell types to avoid cell-type imbalance bias. Trajectory inference was performed using Monocle3 v1.0.0 by learning a principal graph and ordering cells along the graph (Cao et al. 2019). We calculated the pseudotime scores for all cells using "pseudotime". We further determined genes with pseudotime-associated gene expression patterns using "graph_test" with "neighbor_graph = principal_graph" (adjusted $P < 0.05$ and Moran's $I > 0.25$).

Hierarchical dendrogram of lung cell types

To assess the transcriptional relationships between lung cell types, we performed hierarchical clustering based on their gene expression profiles. For each cell type, we computed the average expression across all cells and determined transcriptomic similarity using Spearman's correlation coefficients. The distance matrix was derived by subtracting correlation values

from 1, and hierarchical clustering was performed using the "hclust" function with the "ward.D2" method.

Comparison of cell abundance and transcriptomic similarity between adult and neonatal tree sparrows

We quantified differences in cell-type abundance between adult and neonatal birds and evaluated transcriptomic similarity for each cell type using Spearman's correlation. Statistical significance was assessed using Wilcoxon rank-sum test with false discovery rate correction for multiple comparisons.

Cross-species comparison of lung single-cell datasets

To explore conserved and divergent cell states across species, we compiled published mouse and human lung scRNA-seq datasets (Table S8). Given the congruence in major cell types, AT1s, AT2s, FBs, ECs, and SMCs, we focused our comparative analysis on these cell types.

To account for developmental differences, we categorized samples into two broad life stages: adult (defined as sexually mature individuals) and neonates (defined as the neonatal birds, human infants, and mouse pups). Specifically, single-cell lung data were collected from adult people between 46 and 75 years (Travaglini et al. 2020) and from adult mice between 3 and 30 months (Almanzar et al. 2020). We collected single-cell lung data from human neonates between 9 and 22 weeks (He et al. 2022) and mouse pups between postnatal day 1 and 14 (P1 to P14) (Guo et al. 2019). We acknowledge that our cross-species comparisons contained biases stemming from age matching (variation in growth and development, and age gaps between birds and mammals), which could result in some unaccounted difference between species.

Cross-species cell-type mapping

To investigate the conservation of lung cell types across species, we employed cross-species cell-type assignment and gene module extraction (CAME), a heterogeneous graph neural network approach (Liu et al. 2023), to map cell types between tree sparrows, mice, and humans. This method leverages non-one-to-one homologous gene mapping to enable accurate cell-type characterization across evolutionary species (e.g. zebrafish and human) (Liu et al. 2023). Using annotated mouse and human lung cell types as references (Table S8), we predicted corresponding cell types in tree sparrows, conducting separate analyses for neonatal and adult stages to account for developmental differences. The resulting cell-type mappings were visualized using Sankey plots.

Calculation of transcriptomic heterogeneity

We quantified transcriptomic heterogeneity across cell types and species using transcriptomic entropy measurements. To account for potential sampling bias arising from different sample sizes, we calculated both raw entropy from the original data and unstructured entropy from permutated data, defining the true transcriptomic heterogeneity as the difference between these values (raw entropy minus unstructured entropy), following the method described by Ma et al. (2022). The entropy was calculated as follows:

$$H(X_j) = -\sum_{i=1}^{n} P(x_{ij}) log P(x_{ij})$$

Here, j indexes a gene and $H(X_j)$ denotes its transcriptomic entropy. The continuous expression data of gene j are partitioned into n discrete bins based on expression levels. $P(x_{ij})$ represents the probability of expression values of gene j falling into the i-th bin. To ensure comparable entropy measurements across species, the same expression intervals were used to generate bins for each species. The analysis code is available on GitHub (https://github.com/haoyanioz/Cross-species-single-cell-transcriptomic-analysis-for-lung/).

Transcriptomic divergence across tree sparrows, mice, and humans

To assess evolutionary divergence in gene expression patterns, we analyzed transcriptomic differences between homologous cell types across tree sparrows, mice, and humans. For each species pair, we computed average gene expression profiles for each cell type and determined their similarity using Spearman's correlation coefficient (r), with transcriptomic divergence quantified as $1 - r$. Spearman's correlation has been used here because it is less sensitive to non-normally distributed data and outlier genes.

Cross-species homologous gene conversion

To enable robust cross-species comparisons, we established a gene homology database across tree sparrows, mice, and humans. We first acquired reference gene models for mouse (GRCm39) and human (GRCh38) genomes from Ensembl release 105, along with the tree sparrow genome assembly from the National Genomics Data Centre (https://db.cngb.org/search/project/CNP0006114/). Using OrthoFinder v2.5.4 (Emms and Kelly 2019), we systematically identified homologous gene pairs between species through cluster analysis of gene families. From these results, we extracted one-to-one single-copy orthologs. To ensure high-confidence ortholog assignments, we further refined our gene list by applying a stringent best-to-best reciprocal BLASTp v 2.12.0 + with an E-value cutoff of 1e−10. Finally, we integrated established human-mouse gene homology information from the BioMart database to supplement and validate our orthology predictions, creating a framework for comparative transcriptomic analyses (http://www.ensembl.org/biomart/martview).

Quantify cell-type-specific conserved and divergent gene expression

To identify cell-type-specific conserved and divergent gene expression across species, we focused our analysis exclusively on

one-to-one orthologs shared by tree sparrows, mice, and humans. For each homologous lung cell type, we calculated average gene expression values across all cells within each sample. To minimize technical artifacts arising from cross-species and developmental stage comparisons, we employed gene-rank normalization (Geirsdottir et al. 2019) rather than using raw expression values. In this approach, each orthologous gene within a sample was assigned a normalized rank score (0–1 scale, with a floor of 0.5) based on its expression level relative to other genes in that sample. This rank-based transformation effectively standardized expression measurements across diverse samples. We excluded genes with a score below 0.8 across all samples (i.e. lowly expressed genes).

The rank-transformed data were log-transformed and subjected to PCA using "gmodels" v2.18.1 (Warnes et al. 2022). We then performed hierarchical clustering with complete linkage agglomeration and correlation distance metrics to identify groups of genes and samples with similar expression profiles, visualized through heatmaps. To systematically classify expression patterns, we applied *k*-means clustering using the "kmeans" function to the ranked expression data, with the optimal cluster number (*k*) determined via the elbow method. This approach enabled us to distinguish between evolutionary conserved genes (sharing similar expression patterns across species) and species-specific genes. Functional enrichment analysis of the resulting gene clusters was performed using ClueGO v2.5.10 with the GO biological process database (EBI-UniProt-GOA-ACAP-ARAP_25.05.2022_00h00) (Bindea et al. 2009). We employed a right-sided hypergeometric test with Benjamini–Hochberg correction, applying stringent significance thresholds ($P < 0.001$ for conserved genes and $P < 0.05$ for species-specific genes).

Cell-cycle analysis

To elucidate the functional significance of avian-enriched genes, the full-length coding sequences of tree sparrow *PSCA* (accession XM_039714568.1) was synthesized and directionally cloned into the pcDNA3.1(+) expression vector using *KpnI* restriction sites to generate C-terminally 3×FLAG-tagged fusion proteins. All plasmid constructs were verified by Sanger sequencing, and plasmid DNA was purified using an EndoFree Plasmid Maxi Kit (Tiangan).

Chicken DF-1 fibroblasts, maintained in high-glucose Dulbecco's Modified Eagle Medium (Gibco, Cat. No. 10566016, with GlutaMAX™) supplemented with 10% fetal bovine serum, were transiently transfected with either the *PSCA* expression construct or an empty-vector control using Lipofectamine 2000 reagent (Invitrogen Cat. No. 11668027). Forty-eight hours post-transfection, cells were harvested by trypsinization, fixed in 70% ethanol at −20 °C overnight, and stained with propidium iodide containing RNase A. Cellular DNA content was analyzed by flow cytometry (BD Fortessa), and cell-cycle distributions (G0/G1, S, and G2/M phases) were quantified using ModFit software.

We also performed an EdU proliferation assay. At 48 h post-transfection, cells were incubated with 10 μM 5-ethynyl-2′-deoxyuridine (EdU, Beyotime) for 2 h. After fixation with 4% paraformaldehyde and permeabilization, EdU detection was carried out using Azide 488 following the manufacturer's protocol. Nuclei were counterstained with DAPI. Images were acquired on a Zeiss LSM 880 confocal microscope and analyzed using ImageJ to quantify the percentage of EdU-positive nuclei.

All transfections and subsequent analyses were performed in biological triplicate. Data are presented as the mean ± standard error. Statistical comparisons between the *PSCA* expression construct and the vector control were performed using an unpaired, two-tailed Student's *t*-test.

Cell–cell crosstalk analyses

To investigate potential intercellular communication networks among lung cell types, we performed cell–cell interaction analyses using CellPhoneDB v4.0.0 (Efremova et al. 2020), focusing on known receptor–ligand pairs. For robust detection, we only considered receptor and ligand genes expressed in >10% of cells within a given cell type. Statistically significant interactions ($P < 0.05$) were used to construct cell-type interaction networks, which were visualized using Cytoscape v3.9.1 (Shannon et al. 2003). Through comparative analysis within mouse and human datasets, we specifically identified tree sparrow-enriched cell–cell interactions as those present in tree sparrows but absent in both mammalian species.

FISH

To validate selected ligand–receptor interactions predicted by CellPhoneDB at tissue level, dual-color FISH was performed on lung sections from adult tree sparrows. Lungs were fixed in 4% paraformaldehyde at 4 °C, embedded in paraffin, and sectioned at 5 μm onto RNase-free slides.

Gene-specific antisense RNA probes targeting *APOA1*, *ABCA1*, and *DCN* were generated. Briefly, coding regions were PCR-amplified (primers are listed in Table S9) and used as templates for in vitro transcription with digoxigenin- or fluorescein-labeling kits (Roche), according to the manufacturer's instructions. Dual-color FISH was performed as previously described (Mao et al. 2019). Probe pairs (dig-*APOA1*/flu-*ABCA1*, dig-*APOA1*/flu-*DCN*, and flu-*ABCA1*/dig-*DCN*) were mixed at a 1:1 ratio (final concentration 1 ng/μL per probe) and hybridized to tissue sections. Following hybridization, slides were incubated with horseradish peroxidase-conjugated anti-digoxigenin and anti-fluorescein antibodies. Signals were amplified using the TSA™ Plus Fluorescence System (PerkinElmer); Cyanine3-tyramide was used for digoxigenin probes and fluorescein-tyramide for fluorescein probes. Nuclei were counterstained with DAPI. Fluorescent images were acquired on a Zeiss LSM 880 confocal microscope with a 63× oil-immersion objective with a numerical aperture of 1.4. All acquisition settings were kept consistent across samples to facilitate qualitative comparison of spatial expression patterns.

Identification of viral entry genes

To identify potential viral entry genes in the respiratory system, we systematically compiled a list of known respiratory virus receptor genes from published literature (Ramos and Fernandez-Sesma 2012; Pirooz et al. 2014; Travaglini et al. 2020). We then analyzed the expression patterns of these viral entry-related genes across all lung cell types in tree sparrows, mice, and humans.

Immunofluorescence staining

To validate the expression patterns of key viral entry receptors identified in our transcriptomic analysis, we performed immunofluorescence staining following established protocols (Angueyra et al. 2023). Tissue samples (lungs and intestines) were fixed overnight in 4% paraformaldehyde at 4 °C, followed by dehydration in 30% sucrose solutions. The intestinal samples served as positive controls for *ACE2* expression validation (Fig. 4f). After embedding in OCT compound, 5 μm cryosections were prepared and equilibrated to room temperature for 30 min. Sections were permeabilized with 0.1% Triton X-100 and blocked with 10% goat serum before incubating with primary antibodies against *ACE2* (1:200 dilution) and *EGFR* (1:100 dilution) at 4 °C overnight. Following PBS washes, sections were incubated with Alexa Fluor 488-conjugated secondary antibodies (Thermo Fisher) for 2 h at 37 °C, with nuclear counterstaining using DAPI. Fluorescence signals were visualized and captured using a Zeiss LSM880 laser confocal microscope.

Author contributions

This study was conceived by W.Z., S.Z., P.G.P.E., and Y.Q. Research was supervised by Z.Z., W.Z., F.M., F.L., S.Z., and Y.Q. Data analyses were carried out by Y.H. and X.Z. Experiments were performed by Q.Z. and F.M. Field work was carried out by H.S., L.M., and Y.F. The manuscript was written by Y.Q. with contributions from W.Z., Z.Z., P.G.P.E., and S.Z. All authors read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgements

We thank Kesheng Wang, Yuehan Xue, Nan Fang, Hongjiu Zhang, and Liang Ma for discussions and technical support in cell isolation and library preparation for single-cell sequencing.

Funding

This research was funded by the National Natural Science Foundation of China (NSFCU23A20162 and NSFC32020103005).

Conflicts of interest

None declared.

Data availability

The single-cell data for lung tissue of the tree sparrow that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA) (https://db.cngb.org/cnsa/) of China National GeneBank DataBase (CNGBdb) with accession number CNP0006422.

Code availability

Analysis scripts are available on GitHub (https://github.com/haoyanioz/Cross-species-single-cell-transcriptomic-analysis-for-lung).

References

Ahlmann-Eltze C, Huber W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics*. 2021:36:5701–5702. https://doi.org/10.1093/bioinformatics/btaa1009.

Almanzar N *et al.* A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*. 2020:583:590–595. https://doi.org/10.1038/s41586-020-2496-1.

Angueyra JM *et al.* Transcription factors underlying photoreceptor diversity. *eLife*. 2023:12:e81579. https://doi.org/10.7554/eLife.81579.

Bernhard W *et al.* Pulmonary surfactant in birds: coping with surface tension in a tubular lung. *Am J Physiol Regul Integr Comp Physiol*. 2001:281:R327–R337. https://doi.org/10.1152/ajpregu.2001.281.1.R327.

Bindea G *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009:25:1091–1093. https://doi.org/10.1093/bioinformatics/btp101.

Bjørnstad S, Paulsen RE, Erichsen A, Glover JC, Roald B. Type I and II pneumocyte differentiation in the developing fetal chicken lung: conservation of pivotal proteins from birds to human in the struggle for life at birth. *Neonatology*. 2014:105:112–120. https://doi.org/10.1159/000355346.

Cao J *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019:566:496–502. https://doi.org/10.1038/s41586-019-0969-x.

Cardoso-Moreira M *et al.* Gene expression across mammalian organ development. *Nature*. 2019:571:505–509. https://doi.org/10.1038/s41586-019-1338-5.

Carroll SB. *Endless forms most beautiful: the new science of evo devo and the making of the animal kingdom*. W. W. Norton & Company; 2005.

Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009:37:W305–W311. https://doi.org/10.1093/nar/gkp427.

Chen D *et al.* Single cell atlas for 11 non-model mammals, reptiles and birds. *Nat Commun*. 2021:12:7083. https://doi.org/10.1038/s41467-021-27162-2.

Clementi N *et al.* Viral respiratory pathogens and lung injury. *Clin Microbiol Rev*. 2021:34:e00103-20. https://doi.org/10.1128/CMR.00103-20.

Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature*. 2014:507:190–194. https://doi.org/10.1038/nature12930.

Dobin A *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013:29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dura B *et al.* scFTD-seq: freeze-thaw lysis based, portable approach toward highly distributed single-cell 3′ mRNA

profiling. *Nucleic Acids Res*. 2019:47:e16. https://doi.org/10.1093/nar/gky1173.

Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc*. 2020:15:1484–1506. https://doi.org/10.1038/s41596-020-0292-x.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y.

Geirsdottir L *et al.* Cross-species single-cell analysis reveals divergence of the primate microglia program. *Cell*. 2019:179:1609–-1622.e16. https://doi.org/10.1016/j.cell.2019.11.010.

Gerke V *et al.* Annexins—a family of proteins with distinctive tastes for cell signaling and membrane dynamics. *Nat Commun*. 2024:15:1574. https://doi.org/10.1038/s41467-024-45954-0.

Guo M *et al.* Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat Commun*. 2019:10:37. https://doi.org/10.1038/s41467-018-07770-1.

Hao Y *et al.* Integrated analysis of multimodal single-cell data. *Cell*. 2021:184:3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

Harrington WN, Kackos CM, Webby RJ. The evolution and future of influenza pandemic preparedness. *Exp Mol Med*. 2021:53:737–749. https://doi.org/10.1038/s12276-021-00603-0.

He P *et al.* A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell*. 2022:185:4841–4860.e25. https://doi.org/10.1016/j.cell.2022.11.005.

Hillenius Willem J, Ruben John A. The evolution of endothermy in terrestrial vertebrates: who? When? Why? *Physiol Biochem Zool*. 2004:77:1019–1042. https://doi.org/10.1086/425185.

Ito T *et al.* Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J Virol*. 1998:72:7367–7373. https://doi.org/10.1128/JVI.72.9.7367-7373.1998.

Ivy CM, Scott GR. Control of breathing and the circulation in high-altitude mammals and birds. *Comp Biochem Physiol A Physiol*. 2015:186:66–74. https://doi.org/10.1016/j.cbpa.2014.10.009.

Kagoshima M, Ito T, Hitoshi K, Yoshio G. Diverse gene expression and function of semaphorins in developing lung: positive and negative regulatory roles of semaphorins in lung branching morphogenesis. *Genes Cells*. 2001:6:559–571. https://doi.org/10.1046/j.1365-2443.2001.00441.x.

Kawaguchi T *et al.* Clinical significance of prostate stem cell antigen in human lung cancer. *Cancer Res*. 2008:68:4518–4524. https://doi.org/10.1158/0008-5472.CAN-07-5999.

Kotton DN, Morrisey EE. Lung regeneration: mechanisms, applications and emerging stem cell populations. *Nat Med*. 2014:20:822–832. https://doi.org/10.1038/nm.3642.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014:30:923–930. https://doi.org/10.1093/bioinformatics/btt656.

Liu XY, Shen QL, Zhang SH. Cross-species cell-type assignment from single-cell RNA-seq data by a heterogeneous graph neural network. *Genome Res*. 2023:33:96–111. https://doi.org/10.1101/gr.276868.122.

Loscertales M *et al.* Type IV collagen drives alveolar epithelial–endothelial association and the morphogenetic movements of septation. *BMC Biol*. 2016:14:59. https://doi.org/10.1186/s12915-016-0281-2.

Ma S *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science*. 2022:377:eabo7257. https://doi.org/10.1126/science.abo7257.

Maina JN. A systematic study of the development of the airway (bronchial) system of the avian lung from days 3 to 26 of embryogenesis: a transmission electron microscopic study on the domestic fowl, *Gallus gallus* variant *domesticus*. *Tissue Cell*. 2003:35:375–391. https://doi.org/10.1016/S0040-8166(03)00058-2.

Maina JN. Spectacularly robust! Tensegrity principle explains the mechanical strength of the avian lung. *Respir Physiol Neurobiol*. 2007:155:1–10. https://doi.org/10.1016/j.resp.2006.05.005.

Makanya AN, Djonov V. Parabronchial angioarchitecture in developing and adult chickens. *J Appl Physiol*. 2009:106:1959–1969. https://doi.org/10.1152/japplphysiol.91570.2008.

Mao A, Zhang M, Liu J, Cao Y, Wang Q. PDGF signaling from pharyngeal pouches promotes arch artery morphogenesis. *J Genet Genomics*. 2019:46:551–559. https://doi.org/10.1016/j.jgg.2019.11.004.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011:17:10–12. https://doi.org/10.14806/ej.17.1.200.

McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. 2019:8:329–337.e4. https://doi.org/10.1016/j.cels.2019.03.003.

Monge C, Leon-Velarde F. Physiological adaptation to high altitude: oxygen transport in mammals and birds. *Physiol Rev*. 1991:71:1135–1172. https://doi.org/10.1152/physrev.1991.71.4.1135.

Morrisey EE, Hogan BL. Preparing for the first breath: genetic and cellular mechanisms in lung development. *Dev Cell*. 2010:18:8–23. https://doi.org/10.1016/j.devcel.2009.12.010.

Mund SI, Stampanoni M, Schittny JC. Developmental alveolarization of the mouse lung. *Dev Dyn*. 2008:237:2108–2116. https://doi.org/10.1002/dvdy.21633.

Pirooz SD *et al.* UVRAG is required for virus entry through combinatorial interaction with the class C-Vps complex and SNAREs. *Proc Natl Acad Sci U S A.* 2014:111:2716–2721. https://doi.org/10.1073/pnas.1320629111.

Powell FL, Hopkins SR. Comparative physiology of lung complexity: implications for gas exchange. *Physiology*. 2004:19:55–60. https://doi.org/10.1152/nips.01469.2003.

Ramos I, Fernandez-Sesma A. Cell receptors for influenza A viruses and the innate immune response. *Front Microbiol*. 2012:3:117. https://doi.org/10.3389/fmicb.2012.00117.

Raredon MSB *et al.* Single-cell connectomic analysis of adult mammalian lungs. *Sci Adv*. 2019:5:eaaw3851. https://doi.org/10.1126/sciadv.aaw3851.

Rumschlag-Booms E, Rong L. Influenza A virus entry: implications in virulence and future therapeutics. *Adv Virol*. 2013:2013:121924. https://doi.org/10.1155/2013/121924.

Saito A, Horie M, Nagase T. TGF-β signaling in lung health and disease. *Int J Mol Sci*. 2018:19:2460. https://doi.org/10.3390/ijms19082460.

Schittny JC. Development of the lung. *Cell Tissue Res*. 2017:367:427–444. https://doi.org/10.1007/s00441-016-2545-0.

Shannon P *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003:13:2498–2504. https://doi.org/10.1101/gr.1239303.

Shi JZ *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*. 2020:368:1016–1020. https://doi.org/10.1126/science.abb7015.

Spinelli JB, Haigis MC. The multifaceted contributions of mitochondria to cellular metabolism. *Nat Cell Biol*. 2018:20:745–754. https://doi.org/10.1038/s41556-018-0124-1.

Tournier JM, Goldstein GA, Hall DE, Damsky CH, Basbaum CB. Extracellular matrix proteins regulate morphologic and biochemical properties of tracheal gland serous cells through integrins. *Am J Respir Cell Mol Biol*. 1992:6:461–471. https://doi.org/10.1165/ajrcmb/6.5.461.

Travaglini KJ *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*. 2020:587:619–625. https://doi.org/10.1038/s41586-020-2922-4.

Warnes G, Bolker B, Lumley T, Johnson R. gmodels: various R programming tools for model fitting. https://CRAN.R-project.org/package=gmodels; 2022.

Watanabe-Takano H *et al.* Endothelial cells regulate alveolar morphogenesis by constructing basement membranes acting as a scaffold for myofibroblasts. *Nat Commun*. 2024:15:1622. https://doi.org/10.1038/s41467-024-45910-y.

Wei X *et al.* PSCA and MUC1 in non-small-cell lung cancer as targets of chimeric antigen receptor T cells. *OncoImmunology*. 2017:6:e1284722. https://doi.org/10.1080/2162402X.2017.1284722.

West JB. Comparative physiology of the pulmonary blood-gas barrier: the unique avian solution. *Am J Physiol Regul Integr Comp Physiol*. 2009:297:R1625–R1634. https://doi.org/10.1152/ajpregu.00459.2009.

Whitsett JA, Wert SE, Weaver TE. Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu Rev Med*. 2010:61:105–119. https://doi.org/10.1146/annurev.med.60.041807.123500.

Young MD, Behjati S. Soupx removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience*. 2020:9:giaa151. doi:10.1093/gigascience/giaa151.