bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A developmental program that regulates mammalian organ size offsets evolutionary distance

One Sentence Summary:

A cross-species developmental molecular atlas identified the indicator of lung and animal size beyond evolution

Authors: Yuko Shimamura[1]*, Junichi Tanaka[1]*, Miwako Kakiuchi[2], Hemanta Sarmah[1], Akihiro Miura[1], Youngmin Hwang[1], Anri Sawada[1], Zurab Ninish[1], Kazuhiko Yamada[3], James J. Cai[4†], Munemasa Mori[1†]

Affiliations:

1, Columbia Center for Human Development and Division of Pulmonary, Allergy, Critical Care, Department of Medicine, Columbia University Medical Center, New York, NY, USA

2, Department of Preventive Medicine, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

3, Department of Surgery, Johns Hopkins University, Baltimore, MD, USA

4, Department of Veterinary Integrative Biosciences, Department of Electrical & Computer Engineering Interdisciplinary Graduate Program in Genetics Center for Statistical Bioinformatics, Texas A&M University, College Station, TX, USA

* Equal contribution, respectively

† To whom correspondence should be addressed:

Munemasa Mori†, MD, Ph.D.
Assistant Professor of Medicine,
Columbia Center for Human Development (CCHD),
Pulmonary Allergy & Critical Care Medicine, Department of Medicine,
Columbia University Irving Medical Center
Email: mm4452@cumc.columbia.edu
Tel: 212-305-1731

James J. Cai†, Ph.D.
Associate Professor,
Department of Veterinary Integrative Biosciences,
Department of Electrical & Computer Engineering,
Texas A&M University
Email: jcai@cvm.tamu.edu
Tel: 979-458-5482

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract:

Pigs are evolutionarily more distant from humans than mice, but their physiological organs are closest to humans. The molecular program leading to a more than 1,000-fold increase in organ size in pigs and humans over that of mice across evolution has not been elucidated. We generated large-scale transcriptional landscapes throughout swine lung development. Our cross-species single-cell molecular atlas let us discover swine progenitor identities, stage-specific markers, and a core organ-size regulation program (COSRP), well-conserved in swine and humans but less so in mice. Across eight mammalian species, human COSRP promoters showed higher homologies to evolutionary-distant large animals, including pigs, than evolutionary-close small animals. Our study provides a molecular foundation during swine lung development that unveils animal size regulation conserved in the COSRP promoter, independent of genome-wide evolution. COSRP is a critical paradigm for studying thousands-fold changes in biological sizes in evolution, development, cancer, zoology, respirology, organoids, and biotechnology, particularly human-compatible organ generation.

Main Text:

Evolution changes the size of animals to adapt to the environment(*1*, *2*). Organ size regulation is a fundamental program of the developmental process, however, the molecular program that governs large organ size across evolution remains a mystery(*3–5*). None of the technologies achieved functional human internal organ generation because there has been a considerable knowledge gap in how human-compatible organ sizes develop. The search for organ size regulators in specific organogenesis programs in large animals, which show more than 1,000-fold organ size differences compared to small animals, has the great potential to elucidate the hidden mechanisms of organ size diversification in evolution and advance human-compatible organ size bioengineering in various fields (*3–5*). In this regard, we focused on the development of the pig lungs because its molecular program has been a complete mystery for centuries, and early embryonic development in pigs still lacks organ primordia and has not yet led to a dramatic increase in organismal size compared to mice.

The lungs are one of the most diversified organs during evolution and the most complex organs to replicate. The need for life-sustaining gas exchange through the lungs facilitated environmental adaptation by evolving the lung's developmental program over billions of years, allowing species-specific physiological morphologies, surfactant protein production, and unique circulatory systems(*6*, *7*). Conversely, species such as the plethodontid salamander, which breaths through skins, are lungless by evolutionary design(*8*). Mice and swine have been used as essential model organisms for decades, and both harbor life-sustaining, well-lobulated, branched lungs(*9–13*). Multiple studies of comparative genomic evolution show that pigs primitively separated around 97 million years ago from mice and humans(*14–17*). Mice and humans were divided about six million years later. Thus, swine genome-wide sequences are evolutionarily apart from humans as compared to mice(*14–17*). Contradictorily, pigs harbor physiological organs similar to humans, distinct from small rodents by more than 1,000-fold in size(*17*, *18*). Given the prolific feature and the physiological similarities to humans, pigs are considered an ideal resort for xenogeneic organ transplantation and the host animals for human organ bioengineering by blastocyst complementation for the ultimate cure for various refractory diseases(*19*, *20*). Although many signaling pathways, including Hippo-Yap, JNK, and IGF-1(*3*), are involved in organ size regulation, their signaling alone does not explain the regulatory mechanism for more than 1,000-fold difference in organ size between small animals and animals of comparable size to humans. To address this question, we performed comparative cross-species bioinformatic analyses in pig, mouse, and human lung development.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Pig embryonic lungs develop exponentially along with their body size.

The mouse's developing lung stage has been molecularly and morphologically well-characterized, whereas swine lung development is rarely characterized (*10*, *13*). To fill the huge knowledge gap on the molecular program on how and when swine lung organ size change is initiated, we harvested developing lungs from miniature pigs harboring human-compatible organ size(*19*) (**Fig. 1a**). Based on the gross morphology, we identified embryonic gestation day18 and 19 (hereafter, E18 and E19) as the primary lung bud formation stage (**Fig. 1b**). Interestingly, the swine primary lung bud exhibited slightly enlarged appearance on E19 than E18. After secondary buds appear on E20~E23, sequential pseudostratified epithelial branching occurs (**Fig. 1b**). Through the pseudoglandular stage, the lungs expand their size exponentially from E40 to E70, along with body weight changes (**Fig. 1c**).

Pigs and humans share unique organ size regulator programs, distinct from mice.

Although swine are evolutionarily more distant from humans than mice(*14–17*), the molecular program required for the acquisition of a human-compatible lung organ size phenotype remains elusive (**Fig. 1d**). We isolated RNA from miniature swine lung tissues at E26, E40, and E70 for bulk RNA-seq analysis (**Fig. 1a**) and defined swine developmental stages by gross lung morphology, histology, and cross-species transcriptome (**Fig. 1a, Extended Data Fig. 1, and Supplementary Table 1**). To extract latent variables underlying phenotypic divergence across species, we performed a principal component (PC) analysis. Using PC1 (39% variance) and PC2 (28% variance), the most extensive variant sets in PC analyses confirmed a relatively uniform sampling without significant batch variance (**Extended Data Fig. 2a**). Importantly, PC3 (13% variance) indicates the developmental time course over the species (**Fig. 2a**). With the combination of PC1 and PC3, the most prominent variance among all PCs, we realized that swine developing lungs have a closer transcriptome variance to humans than mice do (**Fig. 2a**). To identify the genes involved in the human and swine transcriptome variance proximity, apart from mice, we ranked genes based on their contribution to the PC1 minus axis. We performed GO term analysis on the top 10% of genes. Of note, we found the program of anatomical structure size regulation (ASSR) (GO: 0090066) (**Fig. 2b and Extended Data Fig. 2b**). Unexpectedly, the neuronal-like developmental program (GO:0007409, 0099536, 0051960, 0010975, 0051963) highly overlays ASSR. Thus, we defined the neuronal and ASSR programs as the **c**ore of the **o**rgan **s**ize **r**egulation **p**rogram (COSRP) that includes 230 genes (**Fig. 2b**, red circled area). Further, the Venn chart illustrated that the COSRP neuronal-like program overlapped with 57.1% of the ASSR program (80 out of 140 ASSR genes) (**Fig. 2c and Extended Data Fig. 2c**).

To map which cell type acquires the unique COSRP features conserved in swine and humans but not in mice, we harvested swine-developing lungs (a total of 83,433 single-cell RNA-seq from 11 independent samples; 7 distinct time points for lung mesenchyme and 4 particular time points for $EPCAM^{+}$ FACS-sorted lung epithelium). We utilized SCGEATOOL(*21*) to perform subclustering annotation with combined datasets across time points, as well as data-driven analyses of gene expression patterns, gene-regulatory networks, and trajectory inference. For cross-species comparison, we combined datasets of swine, human, and mouse-developing lungs and analyzed them in an integrated manner (**Fig. 2d**). The UMAP analyses for the transcriptome profile of swine lung epithelium throughout swine lung development revealed relatively closer progenitor identity with that of human lung epithelium than mice (**Fig. 2d**). In the transcriptome profile of the lung mesenchyme, mice were still far apart from humans overall (**Fig. 2d**). Among 230 COSRP genes in the Venn chart (**Extended Data Fig. 2c**), we discovered that 45 genes were expressed in human and swine lung development but rarely in mice by the cross-species 3D-tSNE scRNA-seq analyses (**Fig. 2e**). Unexpectedly, we also discovered that 17 COSRP genes,

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

including the neuronal-like program and ASSR, were ubiquitously conserved in both EPCAM$^+$ pig and human lung epithelium or mesenchyme but not well-conserved in mice (**Fig. 2e, f, g, and Supplementary Table 2**). These ubiquitous gene expression patterns during swine and human lung development but rarely in mice showed three distinct distributions: first, 3 COSRP genes were exclusively enriched in the EPCAM$^+$ epithelium (*SEMA3C, B3GNT2,* and *B4GALT5*); second, 4 genes exclusively enriched in EPCAM$^-$ mesenchyme (*COL5A2*, *ENAH*, *PRKG1*, and *NR2F1*); third, 10 genes were enriched in both, such as (*SRP54, EFNA5, ANTXR1, ADNP, ID4, TERF2LP, ADAMTS7, CCDC85B, TET1,* and *ZFP36L2*) (**Fig. 2e, f, g, Extended Data Fig. 2c, and Supplementary Table 2**). These results suggested that 45 COSRP genes are the responsible organ size regulator conserved in swine and human lung development but rarely in mice over time.

Unique expression pattern of COSRP genes conserved in swine developing lung epithelium

To uncover which cells express COSRP genes in developing lung epithelium and systematically characterize early lung epithelial progenitor molecular identities, we analyzed scRNA-seq data of swine lung epithelium across four-time points from E18, E19, E40, and E70 (27,287 genes for a total of 9,240 live cells) (**Fig. 3a**). Human lung epithelial progenitors exhibit *SOX2*$^+$*SOX9*$^+$ double-positive distal lung bud progenitors, never seen in the mice at the pseudoglandular stage(*22*). On E18 and E19 of the primary lung bud formation stage, *SOX2* and *SOX9* were indeed co-expressed almost entirely in the lung epithelium (**Fig. 3b,** asterisks), and confocal immunofluorescent (IF) protein expression analyses further confirmed their distribution and SPB and SPC protein expression in the distal lung bud regions (**Fig. 3c,** asterisks, **and Extended Data Fig. 3a-d**). In contrast, at the beginning of the sacculation stage on E70, SOX2 and SOX9 rarely overlapped (**Fig. 3b, c, d,** arrow). On the other hand, SOX2 and SOX9 coincided with E40 lung buds, but the SOX2 expression level was relatively decreased compared to the SOX2$^+$ proximal domain (**Fig. 3b, c, d,** arrowhead).

ScRNA-seq analysis further identified 10 epithelial clusters (**Fig. 3e**). E18 and E19, the primary lung bud formation stage, showed similar profiles but were utterly different from E40 and E70, pseudoglandular and canalicular lung bud stages, respectively. Using classical mouse lung epithelial markers(*23*), the dot plot and feature plot visualized airway and alveolar epithelial cells during swine lung development (**Extended Data Fig. 3a, b**). Although it is not related directly to COSRP, we also discovered novel stage-specific markers such as *NES, SLC16A3, FABP3*, and *SLC2A3* explicitly in the *SOX2*$^+$*SOX9*$^+$ double-positive swine primary lung bud epithelial cells while *Slc16a3, Fabp3*, and *Slc2a3* were partially expressed in E9, the mouse primordial lung stage (**Extended Data Fig. 4a, b**). *SLC22A3*, *GPLD1*, *SHISA2*, and *KCNJ15*, were the markers of pseudoglandular~canalicular specific *SOX9*$^+$ swine lung bud epithelial cells, in which *SLC22A3*, *SHISA2*, and *KCNJ15* were well-preserved in human lung development but not in mice (**Extended Data Fig. 4a, b, c**). Among the 45 COSRP genes, we found 12 genes *(SRP54, B4GALT5, B3GNT2, EFNA5, ANTXR1, ADNP, ID4, TERF2IP, SEMA3, ADAMTS7, CCDC85B, TET1,* and *ZFP36L2*) were moderately or highly expressed in most lung epithelium throughout swine lung development (**Fig. 3f, g, and Supplementary Table 2**). The dot plot analyses showed the spatially distinct expression pattern of COSRP genes in the subclusters (**Fig. 3f**). Briefly, *NR2F1* and *ENAH* were exclusively expressed at the primary lung bud epithelium but rarely later. Interestingly, *B4GALT5, SOCS1, and COL5A2* were highly enriched in the SOX9$^+$ distal lung bud. In contrast, *SEMA3C or B3GNT2 was* relatively increased in the proximal epithelial domain on day40 or day70, respectively. *ID4* was enriched in neuroendocrine cells (NEC) (**Fig. 3f, g, and Supplementary Table 2**).

Identification of COSRP genes ubiquitously expressed in swine developing lung mesenchyme

To reveal the COSRP gene expression in the lineage trajectory and diversification of developing pig lung mesenchyme, scRNA-seq datasets of whole lungs on E18, 19, 23, and 25 and peripheral lung

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tissues on E35, E40, and E70 were analyzed using Seurat and classified the cellular subclusters (**Fig. 4a**). We demonstrated that chondrocyte progenitors emerged at the swine primary lung bud stage, marked by *SOX9*. At the E35 and 40 middle pseudoglandular stage, matrix fibroblast and smooth muscle cell (SMC) clusters appeared (**Fig. 4a**). At E70, late pseudoglandular~canalicular stage, SMC, pericyte, and matrix fibroblast clusters were found (**Fig. 4a**). A dot plot and feature plot analysis of swine lung mesenchymal cells showed stage-specific subclusters and their gene expression profiles well-known in mouse studies (**Fig. 4b, c**). For example, *NOTCH3* and *EBF1* were enriched in the vascular pericyte clusters at the canalicular stage. At the swine lung bud stage, a part of proliferative mesenchyme or matrix fibroblast expressed *FGF10*, a critical mitogen for primary lung bud formation and branching morphogenesis(*24*), *WNT2,* a marker for the lung mesenchyme specifier(*25*), *RSPON2*, a feature of lung bud-specific mesenchyme in human lung development(*26*)(**Fig. 4b, c**). *ACTA2, TAGLN, MYH11,* and *HHIP* were highly expressed in the smooth muscle cell (SMC) clusters at the canalicular stage and myofibroblasts at the pseudoglandular stage(*27*). *ACTA2* and *IGF1-positive* myofibroblasts emerged from E23 and differentiated into *ACTA2*- and *FGF18*-positive smooth muscle cells at E70 (**Fig. 4b, c**). Pseudotime analysis revealed stage-specific genes enriched in the primary lung bud fibroblast-specific stage (*CXADR* and *DSP*), primary to pseudoglandular stage (*HMGA1* and *NASP*) (**Fig. 4d, e**). A matrix fibroblast (MF) labeled by *MFAP4, ELN, MEOX2, TCF21,* and *DCN* has been abundantly enriched in swine and human lung development but rare in mice(*28*). We found *MFAP4*, *ELN*, and *TCF21* gradually increased their expression levels during swine lung development and peaked at E70, most likely by its maturation (**Extended Data Fig. 5a, b**). *TCF21* and *DCN* expressions were conserved across swine and humans but not in mice (**Extended Data Fig. 5b**). We identified novel MF markers through scRNA-seq of swine developing lung mesenchyme. *CAV2*, *RGCC*, and *INMT* were highly expressed in swine MFs. These genes were expressed in human lung mesenchymal tissue but were not abundant in mice (**Extended Data Fig. 5c**). These results suggest that the mesenchymal subcluster of peripheral lung tissue reported in humans is highly conserved in pigs. Among the 45 COSRP genes, 7 *(ERF2IP, COL5A2, ZFP36L2, ENAH, PRKG1, NR2F1,* and *ADNP)* were ubiquitously expressed in the lung mesenchyme throughout swine lung development, but the other COSRP genes showed diverse expression patterns (**Fig. 4c and Supplementary Table 2**).

COSRP gene regulatory networks are conserved in swine and humans but not in mice

To elucidate the significant network of cellular functional programs governing the organ size regulators during lung development over mRNA and protein expression patterns, we performed gene regulatory network (GRN) analyses for the 45 COSRP genes. We used swine, human, and mouse lung developmental scRNA-seq databases at each time point by SCGEATOOL(*21*). Based on the scRNA-seq GRN analysis (scGRN), swine lung developmental scGRN showed multiple highly interacting nodes with connected networks similar to humans (**Extended Data Fig. 6a, b**). The highly-dense scGRN network is conserved in swine and human lung development but sparse in mice.

Together, our analyses showed that swine developmental lung's phenotypic similarity of organ size, cellular phenotype, and scGRN with humans is regulated by COSRP that is not well-conserved in mice. However, it was unclear how organ size regulators represented by COSRP offset the evolutionary distance to humans, in contrast to the mice, which is calculated by each species' overall genomic sequences(*14–17*).

COSRP genomic promoters are better conserved in large animals than in small animals

Since COSRP mRNA expression depends on its promoter activation, we speculated that swine and humans' lung organ size phenotypes might rely on the COSRP promoter genomic sequence homologies against the entire genomic evolution. To explore this possibility, we investigated 15 COSRP

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene promoter regions ($pCOSRP_{15}$) and 11 HK gene promoter regions (pHK_{11}) as controls. To keep the evolutionary significance and avoid the complexity of differential length of promoter analyses across the various species, we analyzed 2kb upstream of TSS: transcription starting site for a human COSRP promoter (pCOSRP) (**Fig. 5a**) (*29*).

Strikingly, we discovered that $pCOSRP_{15}$ aligned sequences showed a significantly higher level of homology between humans and pigs (69.7% ±7.01%) than the homologies between mice and humans (59.1%±7.29%) ($P<0.001$***). Conversely, pHK_{11} did not significantly (ns: non-significant) differ between the homology comparison with pigs (59.4%±8.16%) and mice (57.9% ±8.35%) (**Fig. 5a**). We also checked the COSRP gene coding sequence (CDS) alignments ($_{CDS}COSRP_{15}$). However, $_{CDS}COSRP_{15}$ also did not show a significant homology difference in humans between mice (85.3%±9.57%) and pigs (89.1%±8.41%) nor in HK genes (mouse: 91.9%±2.95%, pigs 93.2%±5.23%). Those results indicate that lung organ size regulation depends on the 2kb upstream of TSS for 15 COSRP genes most likely to activate the COSRP program.

COSRP was found by the cross-species bulk RNA-seq analyses when the lung size exponentially increased along with the organismal size's drastic change (**Fig. 1, 2**). Thus, we hypothesized that pCOSRP homology might be an indicator of lung size but also the organismal size itself among mammalian organisms. Since some organisms were not known the genomic sequence of COSRP or HK promoter homologous regions in the Ensemble, we selected 8 animals (Lion, Pig, Horse, Guinea pig, Golden hamster, Mouse, and Shrew mouse) and used 14 COSRP gene's promoter regions $pCOSRP_{14}$ and 10 HK genes (pHK_{10}) for the homology comparisons (**Fig. 5b**). Strikingly the $pCOSRP_{14}$ genomic homology was well-conserved in mammalian large animals compared to small creatures (**Fig. 5b**). The Pigs (*Sus Scrofa*)'s $pCOSRP_{14}$ homologies (% identity and % Gap) to the humans showed significance to all small animals such as Shrew (*Mus phahari*) (% identity*: p=0.0013,* % Gap: *p=0.0017, respectively*), Mouse (*Mus Musculus*) (*p=0.0006, p=0.005*), and hamster (*Mesocricetus auratus*) (*p=0.0040, p=0.034*), but not to the large animals such as Lion (*Panthera Leo*) (*p>0.05, p>0.05*) or Horse (*Equus Caballus*) (*p>0.05, p>0.05*) (**Fig. 5b**). The pHK_{10} homologies to the humans vary and no significant change between pigs and the other animals (*ns*: non-significant), suggesting $pCOSRP_{14}$ is the unique genomic region highly conserved in large animals but less in small animals.

Discussion

Our analyses unlocked previously unknown COSRP regulatory mechanisms for lung size along with the body size that contradicts the species' entire genomic sequence-based evolutionary distance. Since small animals did not show high $pCOSRP_{14}$ homology to humans, even though they are evolutionarily closer to humans than swine, small species most likely needed to lose the $pCOSRP_{14}$ homologies to make their body size smaller to adapt to the environment billions of years ago after the division of pigs and human beings as a survival strategy during evolution. Our results suggest that large organismal size development requires COSRP in developing epithelial and mesenchymal lung progenitors. Further detailed analyses for the COSRP promoters will be required to elucidate which signaling pathways, such as Hippo-Yap, and JNK, coordinate to activate the $pCOSRP_{14}$. Our results indicate that "species" various phenotypic divergence relies more on the responsible promoter element than an entire genomic sequence that determines genome-wide evolutionary distances (*14–17*). Thus, our findings provide a novel paradigm for cross-species analysis, organismal size regulations, and lung organ size across the species, distinct from the classical evolutionary phylogenic tree studies. The discovery of COSRP and pCOSRP can be utilized for various studies that will lead to a better understanding of zoology, evolution, pulmonary medicine, regeneration, development, cancer biology, and bioengineering, particularly organ bioengineering compatible with humans.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reference:

1. G. Horiguchi, H. Tsukaya, Organ size regulation in plants: Insights from compensation. *Front. Plant Sci.* **2**, 1–6 (2011).

2. B. L. M. Hogan, P. a Kolodziej, Organogenesis: molecular mechanisms of tubulogenesis. *Nat. Rev. Genet.* **3**, 513–23 (2002).

3. A. I. Penzo-Mendez, B. Z. Stanger, Organ-size regulation in mammals. *Cold Spring Harb. Perspect. Biol.* **7** (2015), doi:10.1101/cshperspect.a019240.

4. F. Marongiu, M. Serra, E. Laconi, Development versus Evolution in Cancer Biology. *Trends in Cancer*. **4**, 342–348 (2018).

5. D. M. Bryant, K. E. Mostov, From cells to organs: building polarized tissue. *Nat. Rev. Mol. Cell Biol.* **9**, 887–901 (2008).

6. M. Lambertz, K. Grommes, T. Kohlsdorf, S. F. Perry, Lungs of the first amniotes: Why simple if they can be complex? *Biol. Lett.* **11**, 1–5 (2015).

7. C. B. Daniels, S. Orgeig, Pulmonary surfactant: The key to the evolution of air breathing. *News Physiol. Sci.* **18**, 151–157 (2003).

8. Z. R. Lewis, R. Kerney, J. Hanken, Developmental basis of evolutionary lung loss in plethodontid salamanders. *Sci. Adv.* **8**, eabo6108 (2022).

9. J. M. Flint, The development of the lungs. *Am. J. Anat.* **VI.** (1906), doi:https://doi.org/10.1002/aja.1000060102.

10. D. N. Kotton, E. E. Morrisey, Lung regeneration: Mechanisms, applications and emerging stem cell populations. *Nat. Med.* **20**, 822–832 (2014).

11. J. R. Rock, B. L. M. Hogan, Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell Dev. Biol.* **27**, 493–512 (2011).

12. M. Mori, W. V Cardoso, "Chapter 1 - Lung Progenitor Cell Specification and Morphogenesis" in *The Lung: Development, Aging and the Environment*, R. H. E. B. T.-T. L. (Second E. Pinkerton, Ed. (Academic Press, Boston, 2014; http://www.sciencedirect.com/science/article/pii/B9780127999418000018), pp. 3–9.

13. W. V Cardoso, J. Lü, Regulation of early lung morphogenesis: questions, facts and controversies. *Development*. **133**, 1611–24 (2006).

14. M. A. M. Groenen, A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H. J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. F. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S. Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S. H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. M. A. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapio, B. Dibbits, N. Drou, Z. Q. Du, K. Eversole, J. Fadista, S. Fairley, T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. R. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z. L. Hu, S. J. Humphray, T. Hunt, H. Hornshoøj, J. T. Jeon, P. Jern, M. Jones, J. Jurka, H. Kanamori, R. Kapetanovic, J. Kim, J. H. Kim, K. W. Kim, T. H. Kim, G. Larson, K. Lee, K. T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. T. Nguyen, P. Ni, S. J. Oh, S. Onteru, F. Panitz, E. W. Park, H. S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle,

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher, L. B. Schook, Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. **491**, 393–398 (2012).

15. J. Zhu, F. Chen, L. Luo, W. Wu, J. Dai, J. Zhong, X. Lin, C. Chai, P. Ding, L. Liang, S. Wang, X. Ding, Y. Chen, H. Wang, J. Qiu, F. Wang, C. Sun, Y. Zeng, J. Fang, X. Jiang, P. Liu, G. Tang, X. Qiu, X. Zhang, Y. Ruan, S. Jiang, J. Li, S. Zhu, X. Xu, F. Li, Z. Liu, G. Cao, D. Chen, Single-cell atlas of domestic pig cerebral cortex and hypothalamus. *Sci. Bull.* **66**, 1448–1461 (2021).

16. A. D. Foote, Y. Liu, G. W. C. Thomas, T. Vinař, J. Alföldi, J. Deng, S. Dugan, C. E. Van Elk, M. E. Hunter, V. Joshi, Z. Khan, C. Kovar, S. L. Lee, K. Lindblad-Toh, A. Mancia, R. Nielsen, X. Qin, J. Qu, B. J. Raney, N. Vijay, J. B. W. Wolf, M. W. Hahn, D. M. Muzny, K. C. Worley, M. T. P. Gilbert, R. A. Gibbs, Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).

17. L. Jin, Q. Tang, S. Hu, Z. Chen, X. Zhou, B. Zeng, Y. Wang, M. He, Y. Li, L. Gui, L. Shen, K. Long, J. Ma, X. Wang, Z. Chen, Y. Jiang, G. Tang, L. Zhu, F. Liu, B. Zhang, Z. Huang, G. Li, D. Li, V. N. Gladyshev, J. Yin, Y. Gu, X. Li, M. Li, A pig BodyMap transcriptome reveals diverse tissue physiologies and evolutionary dynamics of transcription. *Nat. Commun.* **12** (2021), doi:10.1038/s41467-021-23560-8.

18. E. Sjöstedt, W. Zhong, L. Fagerberg, M. Karlsson, N. Mitsios, C. Adori, P. Oksvold, F. Edfors, A. Limiszewska, F. Hikmet, J. Huang, Y. Du, L. Lin, Z. Dong, L. Yang, X. Liu, H. Jiang, X. Xu, J. Wang, H. Yang, L. Bolund, A. Mardinoglu, C. Zhang, K. von Feilitzen, C. Lindskog, R. Pontén, Y. Luo, T. Hökfelt, M. Uhlén, J. Mulder, An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science (80-.).* **367**, eaay5947 (2020).

19. M. Sykes, D. H. Sachs, Transplanting organs from pigs to humans. *Sci. Immunol.* **4** (2019), doi:10.1126/sciimmunol.aau6298.

20. J. Wu, H. T. Greely, R. Jaenisch, H. Nakauchi, J. Rossant, J. C. I. Belmonte, Stem cells and interspecies chimaeras. *Nature*. **540**, 51–59 (2016).

21. J. J. Cai, scGEAToolbox: A matlab toolbox for single-cell RNA sequencing data analysis. *Bioinformatics*. **36**, 1948–1949 (2020).

22. M. Z. Nikolić, O. Caritg, Q. Jeng, J. A. Johnson, D. Sun, K. J. Howell, J. L. Brady, U. Laresgoiti, G. Allen, R. Butler, M. Zilbauer, A. Giangreco, E. L. Rawlins, Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *Elife*. **Jun 30**, e26575 (2017).

23. M. E. Ardini-Poleske, R. F. Clark, C. Ansong, J. P. Carson, R. A. Corley, G. H. Deutsch, J. S. Hagood, N. Kaminski, T. J. Mariani, S. S. Potter, G. S. Pryhuber, D. Warburton, J. A. Whitsett, S. M. Palmer, N. Ambalavanan, S. Lin, C. Ljungberg, Z. Bar-Joseph, D. Al Alam, R. Figueroa, S. Fraser, D. Koos, R. Lansford, R. Moats, H. Pollack, W. Shi, A. Shirtz, C. Wigfall, M. Adam, B. Aronow, Y. Du, M. Guo, J. Kitzmiller, A. Perl, A. Potter, J. Snowball, S. Wert, K. Wikenheiser-Brokamp, J. Woods, Y. Xu, C. Chan, C. Hill, J. Kirchner, G. C. Clair, J. Laskin, J. E. McDermott, M. Duparc, N. Gaddis, J. Levy, G. Page, H. Pan, T. Nicola, M. Chan, D. Chhabra, C. Espinoza, G. Bandyopadhyay, T. Bushnell, J. Holden-Wiltse, P. Katzman, R. Misra, C. Frevert, S. Gharib, F. Ahangari, LungMAP: The molecular atlas of lung development program. *Am. J. Physiol. - Lung Cell. Mol. Physiol.* **313**, L733–L740 (2017).

24. F. Chen, T. J. Desai, J. Qian, K. Niederreither, J. Lü, W. V Cardoso, Inhibition of Tgf beta signaling by endogenous retinoic acid is essential for primary lung bud induction. *Development*. **134**, 2969–79 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. A. M. Goss, Y. Tian, T. Tsukiyama, E. D. Cohen, D. Zhou, M. M. Lu, T. P. Yamaguchi, E. E. Morrisey, Wnt2/2b and beta-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev. Cell*. **17**, 290–8 (2009).

26. R. F. C. Hein, J. H. Wu, Y. Tsai, A. Wu, A. J. Miller, E. M. Holloway, T. Frum, A. S. Conchola, E. Szenker-ravi, B. Reversade, K. S. Yan, C. J. Kuo, J. R. Spence, R-SPONDIN2+ Mesenchymal Cells Form the Bud Tip Progenitor Niche During Human Lung Development. *bioRxiv* (2021).

27. S. Ghosh, J. R. Paez-Cortez, K. Boppidi, M. Vasconcelos, M. Roy, W. Cardoso, X. Ai, A. Fine, Activation dynamics and signaling properties of Notch3 receptor in the developing pulmonary artery. *J. Biol. Chem.* **286**, 22678–87 (2011).

28. S. Danopoulos, S. Bhattacharya, T. J. Mariani, D. Al Alam, Transcriptional characterisation of human lung cells identifies novel mesenchymal lineage markers. *Eur. Respir. J.* **55** (2020), doi:10.1183/13993003.00746-2019.

29. S. Lee, I. Kohane, S. Kasif, Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes. *BMC Genomics*. **6** (2005), doi:10.1186/1471-2164-6-168.

30. M. J. L. de Hoon, S. Imoto, J. Nolan, S. Miyano, Open source clustering software. *Bioinformatics*. **20**, 1453–1454 (2004).

31. A. J. Saldanha, Java Treeview—extensible visualization of microarray data. *Bioinformatics*. **20**, 3246–3248 (2004).

32. A. Vienna, R: A language and environment for statistical computing. R Foundation for Statistical Computing, . *R Core Team* (2022).

33. P. Bardou, J. Mariette, F. Escudié, C. Djemiel, C. Klopp, Jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics*. **15**, 1–7 (2014).

34. P. Meylan, R. Dreos, G. Ambrosini, R. Groux, P. Bucher, EPD in 2020: enhanced data visualization and extension to ncRNA promoters. *Nucleic Acids Res.* **48**, 65–69 (2019).

35. G. O. C. Cory, R. Garg, R. Cramer, A. J. Ridley, Phosphorylation of tyrosine 291 enhances the ability of WASp to stimulate actin polymerization and filopodium formation. Wiskott-Aldrich Syndrome protein. *J. Biol. Chem.* **277**, 45115–21 (2002).

Acknowledgments:

We thank the members of the Columbia Center for Human Development for the scientific input, members of the Columbia Genome Center for the transcriptional analysis for RNA-seq, the CCTI flow cytometry core (LSRII: NIH S10RR027050), and Columbia Stem Cell Initiative (CSCI) Flow Cytometry core (FACS Area) for FACS and sorting.

Funding:

NIH-NHLBI 1R01 HL148223-01, DoD PR190557, PR191133 to M. M.

DoD GW190096 to J.J.C.

Author contributions:

Conceptualization: M.M.

Methodology: M.M., Y.S., J.T., J.J.C., M.K.

Investigation: Y.S., J.T., H.S., A.M., Y.H., A.S., Z.N., K.Y., M.M.

bioRxiv preprint doi: https://doi.org/10.1101/2022.10.19.512107; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Visualization: Y.S., J.T., M.K., J.J.C., M.K.

Funding acquisition: M.M., J.J.C.

Project administration: M.M.

Supervision: J.J.C., M.M.

Writing – original draft: Y.S., J.T., J.J.C., M.M.

Competing interests: Authors declare that we have no competing interests.

Data and materials availability: Data available upon request

Supplementary Materials

Materials and Methods

Extended Data Fig. 1 to 6

Supplementary Table 1 to 2

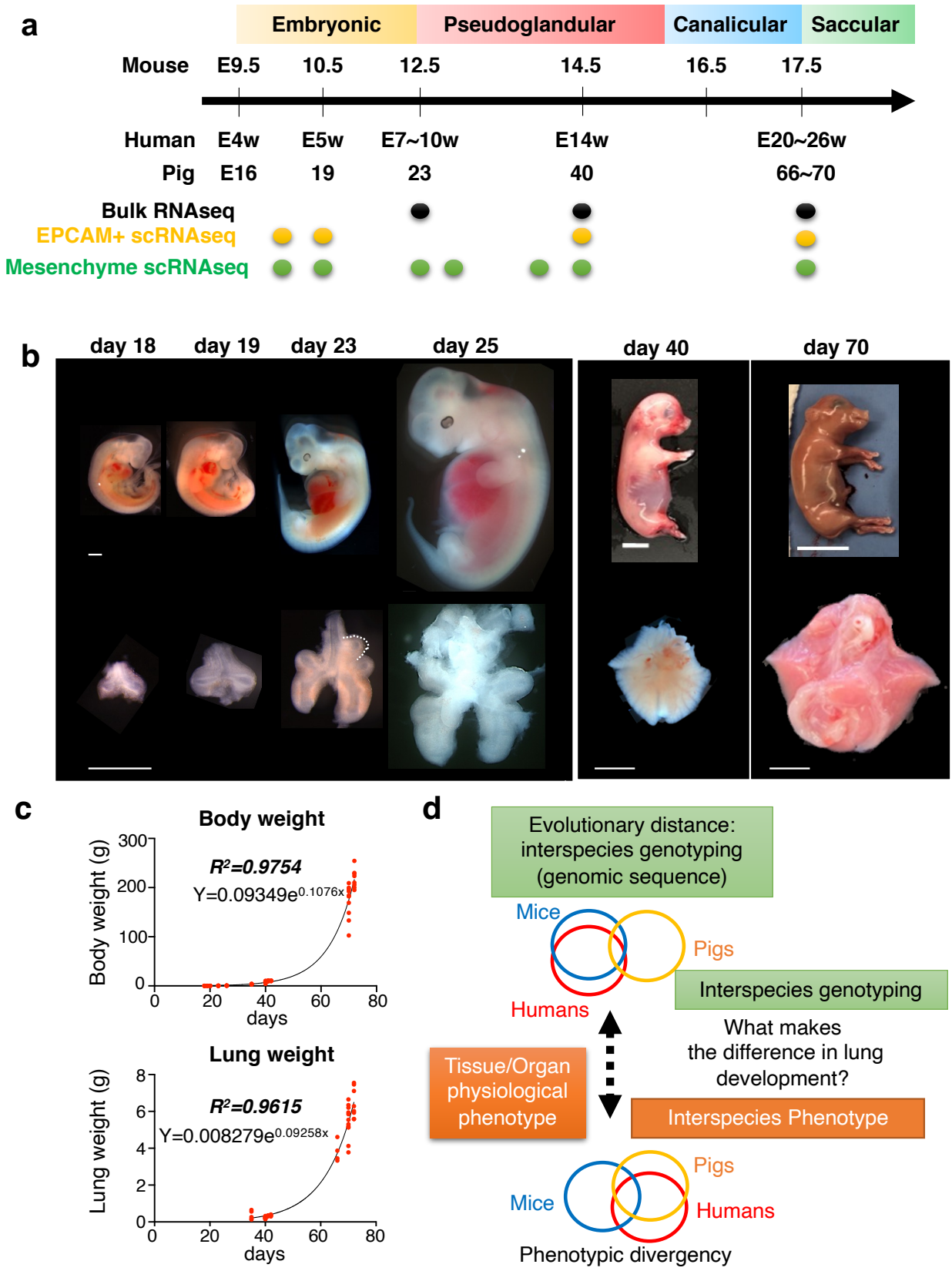
a
Embryonic
Pseudoglandular
Canalicular
Saccular
Mouse
E9.5
10.5
12.5
14.5
16.5
17.5
Human
E4w
E5w
E7~10w
E14w
E20~26w
Pig
E16
19
23
40
66~70
Bulk RNAseq
EPCAM+ scRNAseq
Mesenchyme scRNAseq
b
day 18
day 19
day 23
day 25
day 40
day 70
c
Body weight
R²=0.9754
Y=0.09349e^0.1076x
Body weight (g)
days
Lung weight
R²=0.9615
Y=0.008279e^0.09258x
Lung weight (g)
days
d
Evolutionary distance:
interspecies genotyping
(genomic sequence)
Mice
Pigs
Humans
Interspecies genotyping
What makes
the difference in lung
development?
Tissue/Organ
physiological
phenotype
Interspecies Phenotype
Pigs
Mice
Humans
Phenotypic divergency

Figure 1. The dramatic change of organismal and lung size during swine development (a) Schematic diagram: Developmental stages and sampling of pig lungs compared to mice and humans. Each experiment is indicated by a dot (Black: bulk RNA-seq, Yellow: EPCAM-sorted epithelial scRNA-seq, Green: mesenchyme scRNA-seq). **(b)** The representative gross morphology of pig embryos at each embryonic gestation period (E) 18, 19, 23, 25, 42, 70. **(c)** Graphs: Exponential change of the weight of pig embryos and lungs. Each red dot indicates each embryonic pig specimen. **(d)** The schematic model depicting the discrepancy between interspecies genotyping and phenotype. The human organ size phenotype is closer to a pig than a mouse, while the human genomic sequence is evolutionarily closer to a mouse than a pig. Scale Bars: 1mm (E18-25), 1cm (E40 embryo and lung, and E70 lung), 5cm (E70 lung)

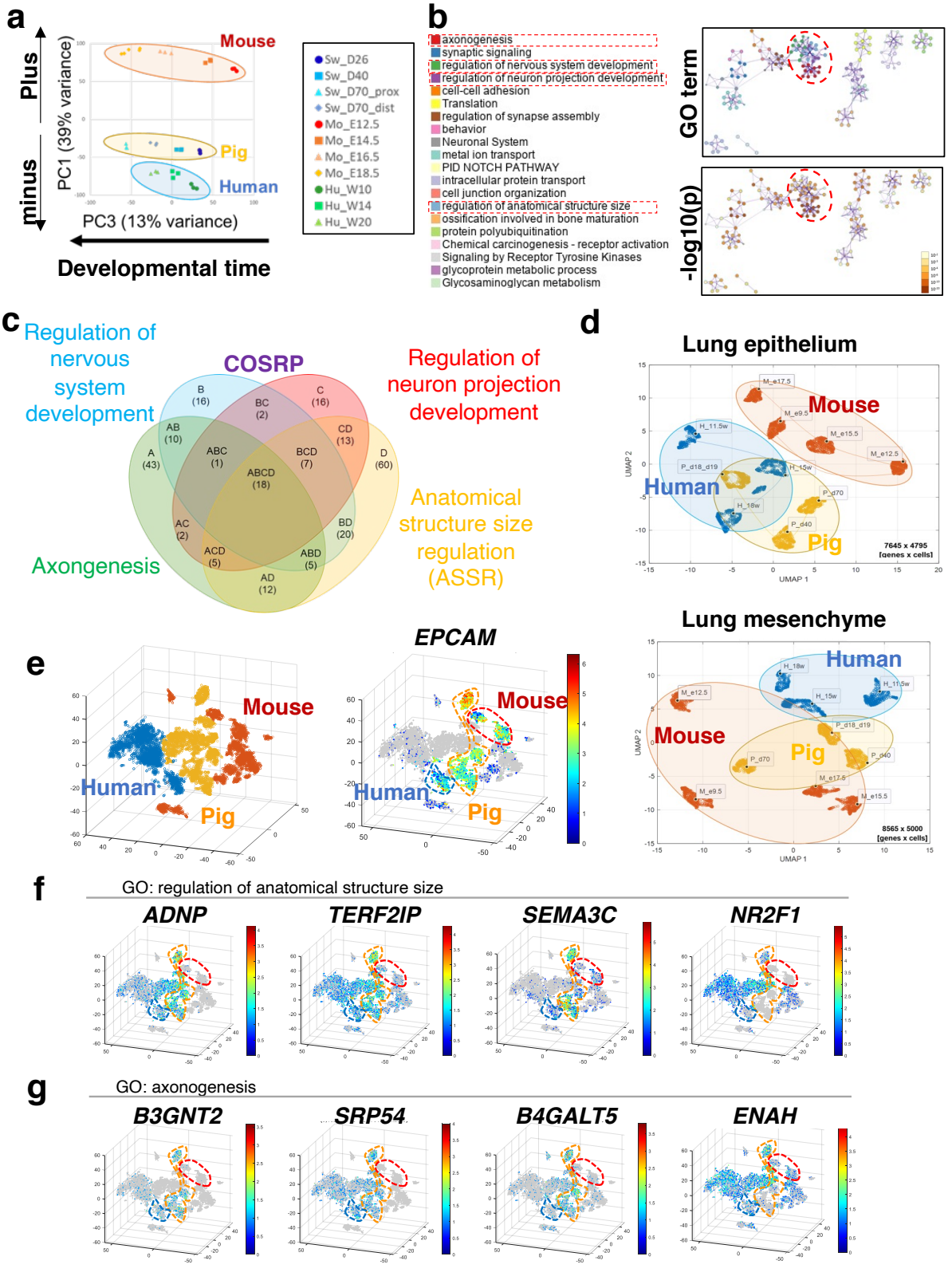
a
Plus
minus
PC1 (39% variance)
PC3 (13% variance)
Developmental time
Mouse
Pig
Human
Sw_D26
Sw_D40
Sw_D70_prox
Sw_D70_dist
Mo_E12.5
Mo_E14.5
Mo_E16.5
Mo_E18.5
Hu_W10
Hu_W14
Hu_W20
b
axonogenesis
synaptic signaling
regulation of nervous system development
regulation of neuron projection development
cell-cell adhesion
Translation
regulation of synapse assembly
behavior
Neuronal System
metal ion transport
PID NOTCH PATHWAY
intracellular protein transport
cell junction organization
regulation of anatomical structure size
ossification involved in bone maturation
protein polyubiquitination
Chemical carcinogenesis - receptor activation
Signaling by Receptor Tyrosine Kinases
glycoprotein metabolic process
Glycosaminoglycan metabolism
GO term
-log10(p)
c
Regulation of nervous system development
COSRP
Regulation of neuron projection development
Anatomical structure size regulation (ASSR)
Axongenesis
B (16)
C (16)
BC (2)
AB (10)
CD (13)
A (43)
ABC (1)
BCD (7)
D (60)
ABCD (18)
AC (2)
BD (20)
ACD (5)
ABD (5)
AD (12)
d
Lung epithelium
Mouse
Human
Pig
UMAP 1
UMAP 2
7645 x 4795 [genes x cells]
Lung mesenchyme
8565 x 5000 [genes x cells]
e
EPCAM
Mouse
Human
Pig
f
GO: regulation of anatomical structure size
ADNP
TERF2IP
SEMA3C
NR2F1
g
GO: axonogenesis
B3GNT2
SRP54
B4GALT5
ENAH

Figure 2. Cross-species transcriptome analyses identify the core of the organ size regulation program (COSRP) during swine lung development. **(a)** PCA analyses showing human, mouse, and pig embryonic lung transcriptional variance. PC3: the developmental time of each species. PC1 minus in the Y axis: the human and swine had similar variance over time, distinct from mice. **(b)** Gene ontology (GO) terms network analysis visualizing the top 10% genes contributing to PC1 minus axis and GO-term associated network: top: Each color indicates GO term category in the left panel. Bottom: colored by significance (P-value: 10^{-2}~10^{-20}). Red dotted circle: COSRP. (**c**) Venn chart plot visualizing 45 COSRP genes composed of four GO terms; Axonogenesis, Regulation of nervous system development, Regulation of neuron projection development, and Anatomical structure size regulation (ASSR). **(d)** Cross-species scRNA-seq UMAP: swine transcriptome closer to humans than mice in developing lung epithelium (top) and mesenchyme (bottom) over time. **(e)** 3D tSNE plot: human, mouse, and swine $EPCAM^+$ lung epithelium and $EPCAM^-$ mesenchyme scRNA-seq combined data. 3D tSNE feature plot for representative COSRP genes: **(f)** *ADNP, TERF2IP, SEMA3C,* and *NR2F1*. **(g)** *B3GNT2, SRP54, B4GALT5,* and *ENAH* were highly enriched in swing and human epithelium or mesenchyme but rarely in mice.

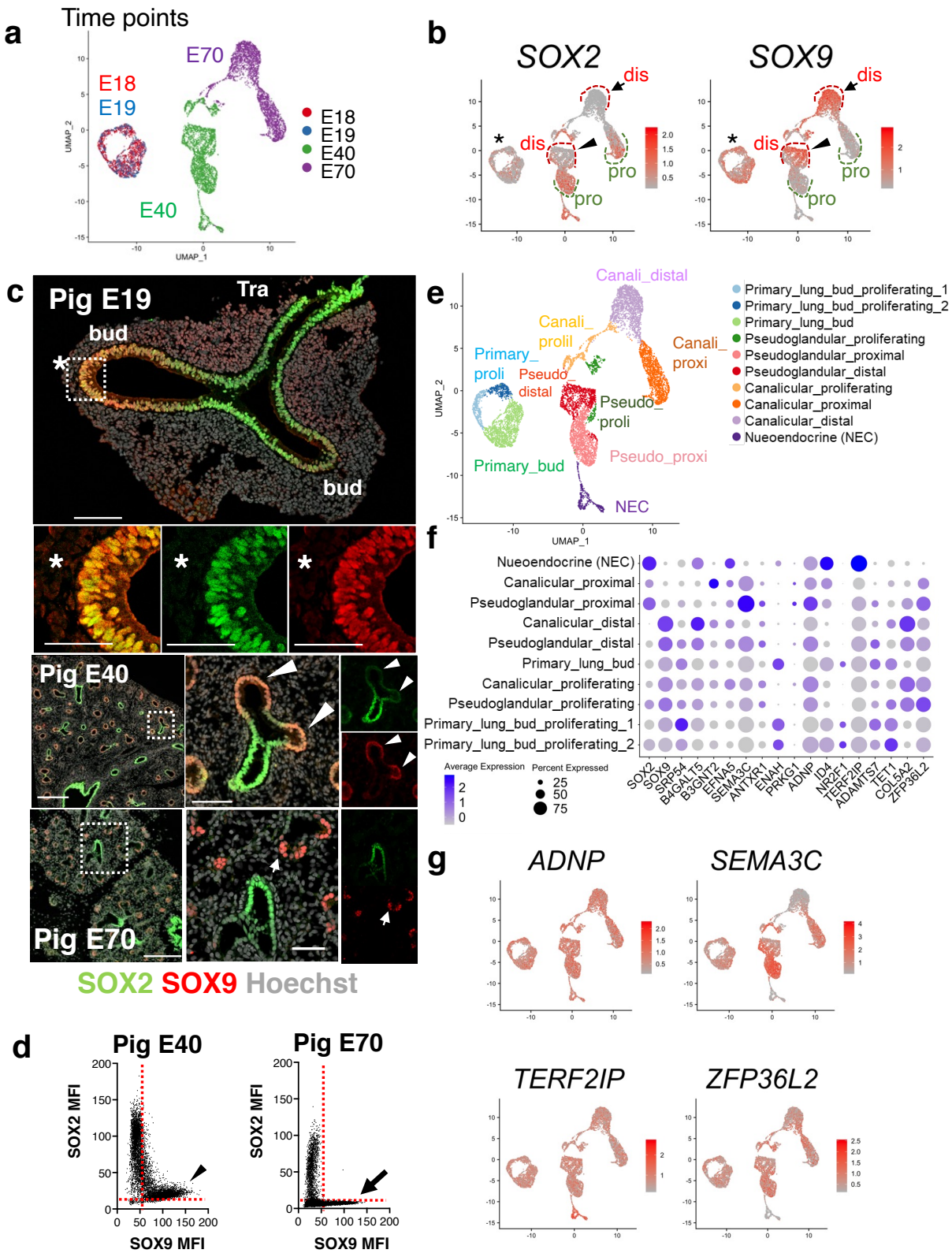

Figure 3. A unique human-like embryonic porcine lung epithelial program. UMAP feature plot for scRNA-seq depicting the samplings from embryonic lung epithelial cells from pig E18, E19, E40, and E70

(a), *SOX2* and *SOX9* gene expression patterns (**b**), ten clusters in swine developing epithelial cells (**e**), and a ubiquitous expression of COSRP genes (ADNP, SEMA3C, *TERF2IP,* and *ZFP36L2)* in swine developing epithelial cells throughout lung development (**g**). **(c)** confocal IF imaging: embryonic pig lungs at E19, E40, and E70: SOX2 (green) and SOX9 (red) were co-expressed on the primary lung buds (asterisks*) and day40 distal lung buds (arrowheads), while only SOX9 were present on the day70 distal lung buds (arrows). **(d)** Graphs: qualitative morphometric analysis for SOX2 and SOX9 protein expression in the nucleus of pig E40 and E70 epithelial cells. Each dot: mean fluorescent intensity (MFI) of Sox2 or Sox9 in a cell nucleus. Arrow; day70 distal epithelial cells, arrowhead: day40 distal epithelial cells. **(f)** Dot plot showing the expression levels of *SOX2, SOX9*, and representative COSRP genes. Dis: distal (red dot lines), Pro: proximal (green dot lines) lung epithelial compartment. Scale bars: top: 100μm, enlarged: 50μm, middle 500μm, enlarged: 100μm, Bottom: 200μm, enlarged: 50μm

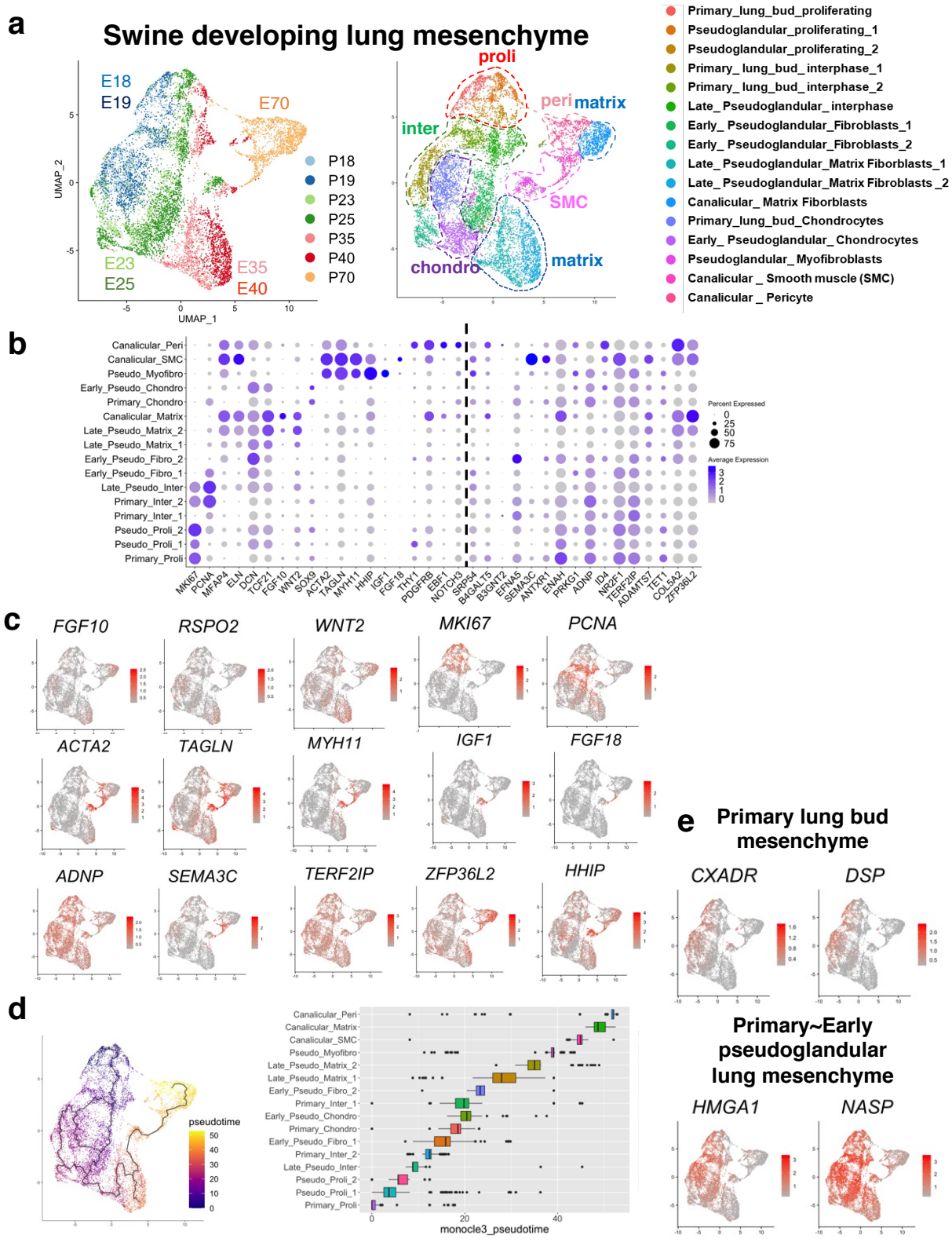
a
Swine developing lung mesenchyme
E18
E19
E70
E23
E25
E35
E40
P18
P19
P23
P25
P35
P40
P70
UMAP_1
UMAP_2
proli
peri
matrix
inter
SMC
chondro
matrix
Primary_lung_bud_proliferating
Pseudoglandular_proliferating_1
Pseudoglandular_proliferating_2
Primary_ lung_bud_ interphase_1
Primary_ lung_bud_ interphase_2
Late_ Pseudoglandular_ interphase
Early_ Pseudoglandular_Fibroblasts_1
Early_ Pseudoglandular_Fibroblasts_2
Late_ Pseudoglandular_Matrix Fiborblasts_1
Late_ Pseudoglandular_Matrix Fibroblasts _2
Canalicular_ Matrix Fiborblasts
Primary_lung_bud_Chondrocytes
Early_ Pseudoglandular_ Chondrocytes
Pseudoglandular_ Myofibroblasts
Canalicular _ Smooth muscle (SMC)
Canalicular _ Pericyte
b
Canalicular_Peri
Canalicular_SMC
Pseudo_Myofibro
Early_Pseudo_Chondro
Primary_Chondro
Canalicular_Matrix
Late_Pseudo_Matrix_2
Late_Pseudo_Matrix_1
Early_Pseudo_Fibro_2
Early_Pseudo_Fibro_1
Late_Pseudo_Inter
Primary_Inter_2
Primary_Inter_1
Pseudo_Proli_2
Pseudo_Proli_1
Primary_Proli
MKI67 PCNA MFAP4 ELN DCN TCF21 FGF10 WNT2 SOX9 ACTA2 TAGLN MYH11 HHIP IGF1 FGF18 THY1 PDGFRB EBF1 NOTCH3 SRPX4 B4GALT5 B3GNT2 EFNA5 SEMA3C ANTXR1 ENAH PRKG1 ADNP ID4 NR2F1 TERF2IP ADAMTS7 TET1 COL5A2 ZFP36L2
Percent Expressed
0
25
50
75
Average Expression
3
2
1
0
c
FGF10
RSPO2
WNT2
MKI67
PCNA
ACTA2
TAGLN
MYH11
IGF1
FGF18
ADNP
SEMA3C
TERF2IP
ZFP36L2
HHIP
d
pseudotime
50
40
30
20
10
0
Canalicular_Peri
Canalicular_Matrix
Canalicular_SMC
Pseudo_Myofibro
Late_Pseudo_Matrix_2
Late_Pseudo_Matrix_1
Early_Pseudo_Fibro_2
Primary_Inter_1
Early_Pseudo_Chondro
Primary_Chondro
Early_Pseudo_Fibro_1
Primary_Inter_2
Late_Pseudo_Inter
Pseudo_Proli_2
Pseudo_Proli_1
Primary_Proli
0
20
40
monocle3_pseudotime
e
Primary lung bud mesenchyme
CXADR
DSP
Primary~Early pseudoglandular lung mesenchyme
HMGA1
NASP

Figure 4. COSRP gene transcriptional profiling of embryonic porcine lung mesenchymal cells. (a) UMAP visualization of lung mesenchymal cells from pig E18, E19, E23, E25, E35, E40, and E70. A dot blot (**b**) and UMAP feature plot (**c, e**) for COSRP genes (*ADNP, ENAH, NR2F1, TERF2IP, COL5A2, SEMA3C, and ZFP36L2*) and representative classical lung mesenchymal markers (distal lung bud markers: *FGF10, RSPO2,* lung mesenchyme specifier*: WNT2,* pericyte: *NOTCH3, EBF1, THY1,* smooth muscle cells (SMC) and myofibroblasts*: ACTA2, TAGLN, MYH11, HHIP, FGF18, IGF1,* chondrocyte: *DCN, SOX9, PDGFRB*, Matrix fibroblast*: MFAP4, ELN, TCF21*, proliferating cells (*MKI67, PCNA*))(**c),** and the stage-specific markers: primary lung bud stage (*CXADR* and *DSP*) and primary to pseudoglandular stage (*HMGA1* and *NASP*) (**e**). **(d)** Monocle pseudotime analysis of embryonic porcine lung mesenchymal cells. Cells are labeled by pseudotime. Boxplot showing the distribution of pseudotime within each cell group.

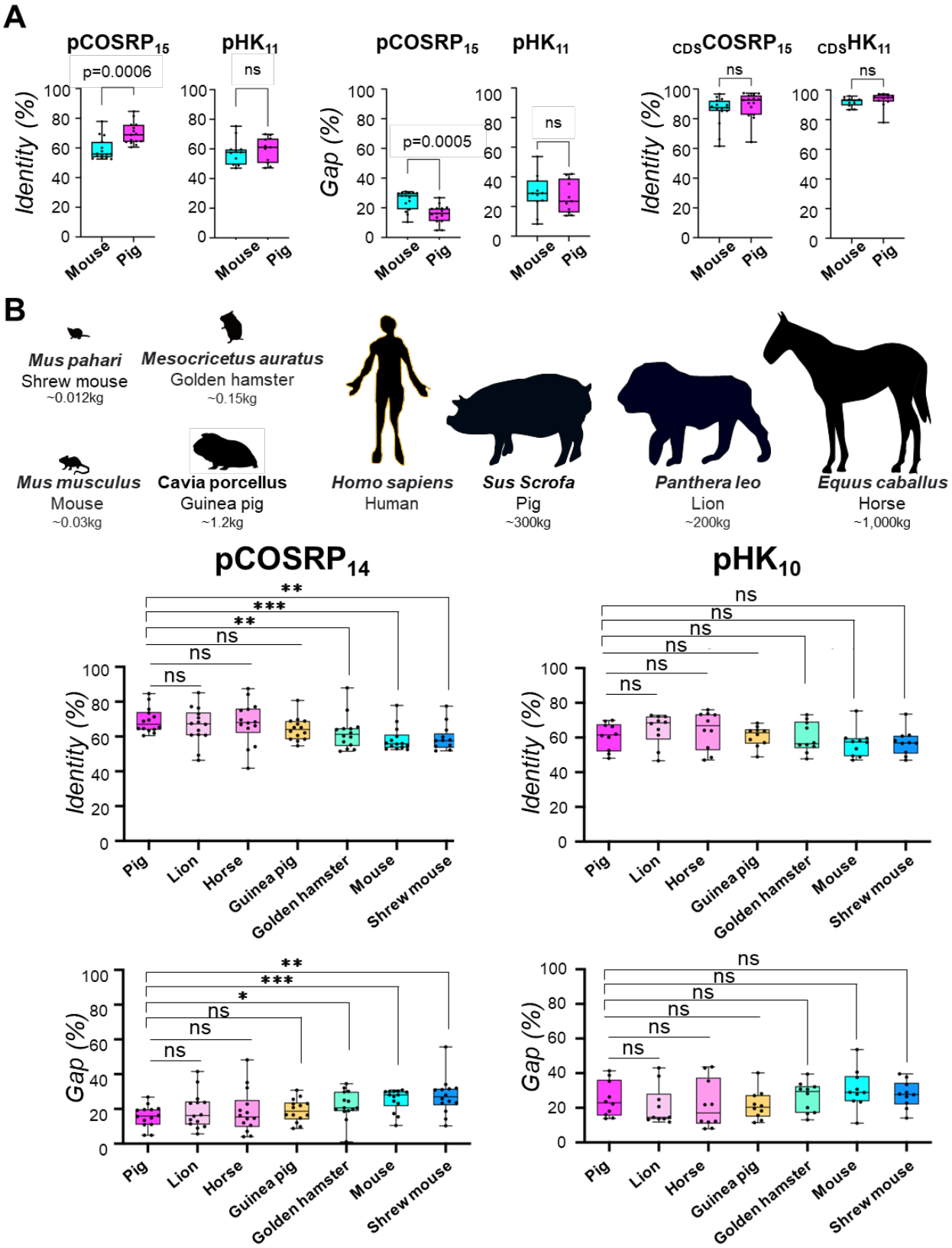
A
pCOSRP15
pHK11
pCOSRP15
pHK11
CDSCOSRP15
CDSHK11
Identity (%)
Gap (%)
Identity (%)
p=0.0006
ns
p=0.0005
ns
ns
ns
Mouse
Pig
B
Mus pahari
Shrew mouse
~0.012kg
Mesocricetus auratus
Golden hamster
~0.15kg
Mus musculus
Mouse
~0.03kg
Cavia porcellus
Guinea pig
~1.2kg
Homo sapiens
Human
Sus Scrofa
Pig
~300kg
Panthera leo
Lion
~200kg
Equus caballus
Horse
~1,000kg
pCOSRP14
pHK10
Identity (%)
Gap (%)
**
***
*
ns
Pig
Lion
Horse
Guinea pig
Golden hamster
Mouse
Shrew mouse

Figure 5. COSRP promoter region encodes a relative organismal size. (a) graphs: A box and whisker plot: COSRP or housekeeping (HK) gene promoter homologies (left: % identity, middle: % gap) of mice and pigs to humans. Right: The Coding sequence (CDS) homologies (% identity) of mice and pigs to humans. The homology of the human COSRP promoter region was significantly lower in mice than in swine ($p=0.0006$, ***$P<0.001$) while not significant in HK promoter regions or CDS or HK regions. (COSRP gene; n=15, HK genes; n=11) **(b)** top: Schematic diagram of large and small animals with representative body weight. Bottom: Graphs: Homologies (% identity, % gap) of each species COSRP, housekeeping (HK) gene promoters to humans. A dot in each graph (**a, b**) represents each COSRP or HK gene in each species. mean±s.e.m. in 14 COSRP genes or 10 HK genes per group. two-tails Student's t-test, *$P<0.05$, **$P<0.01$, ***$P<0.001$, ****$P<0.0001$, ns: statistically non-significant.