

Systematic identification of cell fate regulatory programs using a single-cell atlas of mouse development

Lijiang Fei
Haide Chen
Lifeng Ma
Weigao E
Renying Wang
Xing Fang
Ziming Zhou
Huiyu Sun
Jingjing Wang
Mengmeng Jiang
Xinru Wang
Chengxuan Yu
Yuqing Mei
Danmei Jia
Tingyue Zhang
Xiaoping Han (✉ xhan@zju.edu.cn)

Zhejiang University

Guoji Guo (✉ ggj@zju.edu.cn)
Zhejiang University <https://orcid.org/0000-0002-1716-4621>

Research Article

Keywords: single-cell, cell type, Xbp1, cell fate

Posted Date: April 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1544600/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Systematic identification of cell fate regulatory programs using a single-cell atlas of mouse development

Lijiang Fei^{1†}, Haide Chen^{1,2†}, Lifeng Ma^{1†}, Weigao E^{1†}, Renying Wang^{1†}, Xing Fang^{1,4†}, Ziming Zhou^{1†}, Huiyu Sun¹, Jingjing Wang², Mengmeng Jiang², Xinru Wang¹, Chengxuan Yu¹, Yuqing Mei¹, Danmei Jia¹, Tingyue Zhang², Xiaoping Han^{1*}, Guoji Guo^{1,2,3,4,5*}

¹Center for Stem Cell and Regenerative Medicine, and Bone Marrow Transplantation Center of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310000, China.

²Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou, Zhejiang 311121, China.

³Zhejiang University-University of Edinburgh Institute, Zhejiang University School of Medicine, Zhejiang University, Hangzhou 314400, China.

⁴Zhejiang Provincial Key Lab for Tissue Engineering and Regenerative Medicine, Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cell and Regenerative Medicine, Hangzhou, Zhejiang 310058, China.

⁵Institute of Hematology, Zhejiang University, Hangzhou, Zhejiang 310058, China.

†These authors contributed equally to this work.

*Correspondence author. Email: ggj@zju.edu.cn (G.G.); xhan@zju.edu.cn (X.H.).

Abstract

Waddington's epigenetic landscape is an abstract metaphor frequently used to explain cell fate decisions. Recent advances in single-cell genomics are altering our understanding of the Waddington landscape. Yet, the molecular regulations behind remain poorly understood. We construct a dynamic cell landscape of mouse lineage differentiation at the single-cell level and thereby reveal both lineage-common and lineage-specific regulatory programs during cell type maturation. We verify lineage-common regulatory programs that are universal during the development of invertebrates and vertebrates. In particular, we identify *Xbp1* as an evolutionarily conserved regulator of cell fate determinations across different species. We demonstrate that *Xbp1* transcriptional regulation is important for the stabilization of the genetic network for a wide range of cell types. Our results offer genetic and molecular insights into the gene regulatory programs systematically and provide resources to advance the understanding of the cell fate decisions.

Introduction

The high reproducibility of the developmental process for multicellular organisms suggests a dedicated regulatory program that governs the trajectories of cell fate decisions¹⁻³. According to Waddington's epigenetic landscape theory, various cell types arise from an unstable stem/progenitor state and eventually fall into stable cell fate attractors⁴. The emerging concept of the state manifold derived from single cell data further enhanced our understanding of lineage progression^{5,6}. What are the gene regulatory programs behind the state manifolds? How are they regulated? These are two central questions puzzling the field.

Transcription factors (TF) and gene regulatory networks (GRNs) are known to govern cell fate decisions⁷. *GATA1/PU.1* system makes the binary choice in the process of hematopoietic differentiation⁸. *MyoD* system has critical roles in myogenic cell-lineage specification during development and trans-differentiation⁹. *Oct4-Cdx2* makes the decisions between inner cell mass and trophectoderm cells¹⁰. These studies clarify the importance of lineage-specific transcriptional regulations in different systems. However, the focused analyses on specific regulatory network modules can't show the complete topological structures of a genome scale high-dimensional GRNs.

With breakthroughs in single-cell RNA sequencing (scRNA-seq), single-cell atlases of various developmental stages have been profiled at the organismal level¹¹⁻¹⁶. Single-cell datasets offer unprecedented opportunities to systematically unravel the nature of cell fate regulatory programs^{17,18}. A systematic and global view of multi-lineage, multi-species cell fate gene regulatory modules may help us to understand the lineage specification and maturation. In this study, we determined the molecular content of lineage-common and lineage-specific regulatory programs through multilineage and cross-species analysis. We constructed a time-series mouse cell differentiation atlas (MCDA) to reveal the gene regulatory networks that govern cell fate decisions (Fig. 1a). We characterized a general feature of decreased entropy with less complexity in most lineages along with development. Through cross-species analysis, we summarized conserved features of cellular differentiation, one of which was that ribosomal genes are universally expressed at high levels in stem/progenitor cells. Importantly, we verified *Xbp1* as a lineage-common master regulator that was involved in core fate-determining circuits.

64 Results

65 Construction of Mouse Cell Differentiation Atlas.

66 We performed single-cell transcriptomic analysis on mice at seven life stages ranging from
 67 the early embryonic stage to the mature adult stage: embryonic day (E) 10.5, E12.5, E14.5,
 68 postnatal day (P) 0, P10, P21, and adult. Altogether, we profiled more than 520,000 single cells
 69 (Fig. 1b and Supplementary Table 1-3). The profiled organs, including the brain, heart,
 70 intestine, kidneys, liver, lungs, pancreas, stomach, testes, and uterus, spanned diverse systems.
 71 Previously published E14.5 and adult data^{11,14}, represented approximately 30% of the cells in
 72 the entire dataset. Systemic mouse single-cell atlases of P0, P10, and P21 have not been
 73 depicted thus far. We projected all single cells on a t-distributed stochastic neighbor embedding
 74 (t-SNE) plot and obtained 95 transcriptionally distinct cell populations (Fig. 1c and
 75 Supplementary Table 4). Clusters that were composed of multiple tissues included immune
 76 cells (C9, C16, C18 C25, C29, C34), stromal cells (C13, C20, C22, C26, C28), muscle cells
 77 (C31), and endothelial cells (C8), while epithelial cells differed across tissues and formed
 78 separate clusters (C17, C19, C39) (Extended Data Fig. 1a-c, and Supplementary Table 5).
 79 Moreover, the clusters were arranged in chronological order, showing projections from fetal
 80 progenitors toward adult mature cell types (Fig. 1d). Analysis of differentially expressed genes
 81 in neighboring stages for each tissue showed that the critical period of tissue maturity varied
 82 among different stages. The transition from the E14.5 to P0 stage led to dramatic changes
 83 during development (Extended Data Fig. 1d). Changes from P0 to P10 were dominated by
 84 energy metabolism on account of the different energy sources¹⁹, while changes from P10 to
 85 adulthood focused on pathways of response, transport, and metabolism (Extended Data Fig. 1e
 86 and Supplementary Table 6). We observe lots of distinct clusters from P0 and P10 samples,
 87 indicating continuing cellular transitions after birth. Together, these data represent, to our
 88 knowledge, the systematic single-cell atlas of the complete mouse lineage differentiation from
 89 development to maturation. We also provide an interactive website <http://bis.zju.edu.cn/MCA/>
 90 to enable public access to the data.

91

92 Cellular changes during the mouse development.

93 The tissue effect gave rise to 31.9% of the global variance, which is much more than variance
 94 from the stage and sex effects (Extended Data Fig. 1f, g). We studied dynamic changes in the
 95 kidneys as a representative. After analyzing kidney samples from the E10.5 to adult stages, we
 96 defined 30 clusters with canonical markers^{20,21} that included stromal cells, nephron epithelial
 97 cells, fenestrated endothelial cells, and immune cells (Fig. 2a, Extended Data Fig. 1h, and
 98 Supplementary Table 7). Cells from diverse developmental stages of nephrogenesis were well
 99 captured in our single-cell data, with ureteric bud (UB) cells (*Ret*+, *Gata3*+,), nephron
 100 progenitors cells (NPCs, *Cited1*+, *Gdnf*+, *Six2*+,), proximal s-shaped body (SSB) cells (*Lsp1*+,
 101 *Tmem100*+,), distal SSB cells (*Lhx1*+,), podocytes (Podos, *Podxl*+,), five types of proximal
 102 tubule (PT) cells, ascending and descending loop of Henle (ALOH and DLOH) cells,
 103 connecting nephron tubule (CNT) cells, distal collecting tubule (DCT) cells, two subsets of
 104 intercalated cells (ICs), and principal cells (PCs). Notably, UB cells and NPCs included cells

at the P0 stage, while distal and proximal SSB cells included cells at the P10 stage (Fig. 2b). This result indicated that nephrogenesis continued postnatally instead of being completed before birth in the mice. Moreover, the maturation of renal function continues until the adult stage with gradual physiological changes (Extended Data Fig. 1i).

We then performed cross-tissue single-cell analysis at different stage (Extended Data Fig. 2, Extended Data Fig. 3, and Supplementary Table 8-59) and uncovered 37 uncommon cell populations with interesting gene expression patterns with regard to mouse development (Supplementary Table 60). For example, several cell types were found for co-expressing markers of two cell types. We identified cells that co-expressed makers of myocytes (*Myl9*, *Acta2*) and endothelial cells (*Esam*, *Gng11*) in both brain and intestine at the P0 stage (Fig. 2c, d). The co-immunofluorescence of *Myl9* and *Esam* further confirmed the scRNA-seq results (Fig. 2e, f). Myoendothelial cells may be endowed with multilineage potential like human myoendothelial cells as reported²². In the P10 lung, we verified a special club cell type (*Scgb1a1*, *Scgb3a1*) expressing goblet cell markers (*Tff2*, *Muc5b*) which may be an intermediate cell type during airway epithelial differentiation (Extended Data Fig. 4a, b). In addition, some tissue-specific markers showed ectopic expression in other tissues. For example, we discovered hepatocyte-like cells (*Afp*, *Alb*) in the pancreas at the both P0 and P10 stages, and immunofluorescence assays confirmed their existence (Extended Data Fig. 4c, d). They displayed different expression patterns from liver hepatocytes, and showed high expression of early hepatic stem or progenitor marker *Hnf4a*^{23,24} (Extended Data Fig. 4e, f). Together, progenitor pools with co-expression or ectopic expression patterns may be widely present in developing organs, suggesting the stochasticity of the mammalian state manifolds.

Characterization of regulatory programs in MCDA.

High-resolution MCDA offers a powerful resource for studying the molecular content of manifold attractors through various lineages. To reveal organism-wide characteristics, we applied different potency models based on entropy to qualify the state manifold landscape²⁵⁻²⁹. Entropy decreased continuously along with organ maturation in the most assayed lineages using different computational methods (Fig. 3a, b and Extended Data Fig. 5a-d), revealing a decrease in transcriptional plasticity and an increase in transcriptional stability. Cell type maturation appears to be an event associated with more singular transcriptomes and biological processes.

Cell types represent high- dimensional attractor states of gene regulatory networks (GRNs)³⁰. Transcription factors (TFs) function as important regulators in GRNs to specify cell types and differentiation patterns³¹. To identify critical TFs of cell identity, we took the advantages of both data-driven (SCENIC)³² and database-derived (VIPER-DOROTHEA)³³ methods to estimate the activities of TFs. We achieved over 75% sensitivity based on single-cell datasets (Extended Data Fig. 5e). Over 900 TFs were identified with confidence levels ranging from A (high confidence) to C (low confidence) (Supplementary Table 61). Aggregated heatmaps were constructed to display the specific and common relationships of the TFs and their enriched lineages during development (Fig. 3c and Extended Data Fig. 5f). The neural lineage was characterized by *Dlx1*, *Pou3f3*, and *Sox10*. The *Cebpa* and interferon regulatory transcription factor (IRF) genes marked the immune lineage, while the endothelial

lineage exhibited prominent *Sox17* and *Sox18* expression. Strikingly, hierarchical clustering analysis showed two modules of lineage-sharing TFs (modules 14 and module 15) that were enriched in adult tissues and fetal tissues respectively (Fig. 3c and Extended Data Fig. 5g). Enrichment and occupancy of Hox and zinc-finger families in fetal tissues have previously been associated with embryonic development^{34,35}. The ubiquitous expression cluster in adult tissues was shared for a wide range of lineages, with extensive representation of *Xbp1*, genes of activator protein-1 (AP-1) family, and other molecules. Only 78/268 TFs in this adult multilineage cluster were housekeeping genes³⁶ (Extended Data Fig. 5h). Jun and Fos gene families can dimerize and form AP-1, which has been reported to act as regulators in the differentiation of various cell types^{37,38}. In addition, AP-1 family members have been recently suggested to act as central guardians of somatic cell fate^{39,40}. These highlighted the important roles of AP-1 family members in cell type differentiation and cell identity maintenance. Moreover, these TFs exhibited increasingly upregulated gene expression levels during lineage maturation (Extended Data Fig. 5i), which coincided with decreased entropy in most lineages (Fig. 3a, b and Extended Data Fig. 5a-d). Together, we suspect that these lineage-common TFs function as vital common regulators during cell-type maturation in mouse.

Global features during cell fate decisions across species.

Given that the suite of regulatory genes that control development is ancient⁴¹, we wondered whether GRNs are conserved in invertebrates and vertebrates. We decided to investigate the lineage-specific and lineage-common regulatory elements during evolution. First, we performed a comparative analysis of gene regulation during development in seven species with varying evolutionary distances at single-cell resolution. Development atlases of four invertebrates and three vertebrates were involved, including *Schmidtea mediterranea*¹², *Caenorhabditis elegans*¹⁵, *Ciona intestinalis*¹⁶, *Hydra vulgaris*¹³, *Danio rerio*⁴², *Mus musculus*¹¹, and *Homo sapiens*¹⁴. More than 1,100,000 cells were collected and categorized into 665 cell type pairs for relatively differentiated states and undifferentiated states (Extended Data Fig. 6a and Supplementary Table 62). Partition-based graph abstraction (PAGA)⁴³ was applied to map cell types along the developmental branch for invertebrates (Extended Data Fig. 6b-d). For vertebrates, to minimize the impact of tissue effects, we connected cell states of the same tissue across time based on gene expression similarity⁴⁴, and cell hierarchies of the human lung were shown as an example (Extended Data Fig. 6e).

To explore the common changes in cross-species development, we performed entropy analysis and found that entropy decreased in all seven species along with development, which suggested that the increase in transcriptional stability was evolutionarily conserved (Fig. 4a-f and Extended Data Fig. 7a, b). For the molecular changes, we performed differential gene expression analysis between corresponding cell types pairs and mapped homologous genes to the most advanced human gene symbols to find commonly regulated genes in multiple species (Extended Data Fig. 7c, d, and Supplementary Table 63-64). For all species, the numbers of conserved downregulated genes were greater than those of conserved upregulated genes, which suggests that stem/progenitors have more convergent expression patterns than differentiated cell types⁴⁵ (Extended Data Fig. 7e). Both commonly downregulated and upregulated genes (in at least three species) tended to have more protein-protein interactions (PPIs) than other genes, which indicated that the common regulators were evolutionarily older⁴⁶ (Extended Data Fig. 7f, g). The genes downregulated during development were enriched with ribosomal protein

genes, mitochondrial ribosomal protein genes, and small nuclear ribonucleoprotein genes (Fig. 4g, Extended Data Fig. 7h, and Supplementary Table 65). Notably, *Myc* and *Mycn*, as regulators of ribosome biogenesis^{47,48}, showed high activity scores in the early stages of mouse development (Fig. 4h). They were classified in the common (fetal) module (Supplementary Table 61). These findings were highly consistent with recent studies, which reported that ribosomal protein genes are specific network hubs as robust markers of differentiation potency^{6,49}. In our cross-species entropy analysis, the conserved driving genes in more potent cells were also enriched in ribosomal biogenesis⁵⁰ (Fig. 4i). Suppression patterns of ribosomal protein genes are also discovered in the differentiation of zebrafish hematopoiesis^{51,52}. Ribosomal protein genes appear to be a conserved feature of stemness; they are downregulated during cell-type differentiation. On the other hand, the upregulated genes were highly enriched for immunity pathways (Extended Data Fig. 7i, j, and Supplementary Table 66), which was consistent with recent reports on human and mouse adult tissues^{14,53}. Together, we present a catalog of common features during lineage development from invertebrates to vertebrates.

Gene regulation networks of cell fate decisions across species.

To search for lineage-specific regulators among different species, we systematically aligned cell lineage homologs from each species across massive evolutionary distances. Two strategies were used and compared with high thresholds to infer the reliable and plausible matches. 47/60 cell lineages from seven species were characterized into eight meta-lineages (Extended Data Fig. 8a). The UMAP embedding based on pseudo-bulk cells per species proved the rationality of meta-lineages (Extended Data Fig. 8b, c). Then, the specificity of TFs was characterized with the modified regulon specific scores per species^{54,55}. Lineage-specific TFs displayed sequence similarity within the meta-lineage across species (Extended Data Fig. 8d-j). Vertebrates tended to have more conserved species-specific TFs than invertebrates.

For lineage-common regulators among different species, we found that several commonly upregulated TFs exhibited remarkable convergence, including *XBPI*, *JUND*, *FOSB*, *JUN*, *BHLHE40*, and others (Fig. 5a), consistent with the enriched TFs in various adult mouse tissues (Fig. 3c, Supplementary Table 61). These TFs also displayed strong negative correlations with TFs that were enriched in lineage-specific progenitor cells (*GATA1*, *PAX6*, *NKX6-2*, *NEUROD1*, *SOX10*, *OLIGO2*) in the Human Cell Landscape (HCL) (Fig. 5b, Supplementary Table 67-68). We suspect that these TFs may function as evolutionarily conserved regulators to guide multilineage cells to differentiation and maturity. We found that only one TF, *Xbp1*, stands out in all seven species (Fig. 5a and Extended Data Fig. 7d). Therefore, we attempt to unravel the role of *Xbp1* in cell type maturation. Previous work has emphasized functions of the basic helix-loop-helix (bHLH) transcription factor *Xbp1* for cell differentiation in various cell types including secretory cells⁵⁶, plasma cells⁵⁷, T cells⁵⁸, neurons⁵⁹, hepatocytes⁶⁰, etc. As a putative common regulator, *Xbp1* showed an upregulated expression pattern in most lineages of MCDA (Fig. 5c). We further dissected its regulatory role from a global perspective and found that stem regulators such as *SOX4*, *SON*, and *HES1* are the most negatively correlated with *XBPI* in the HCL (Fig. 5d). In addition, the *XBPI* binding motif in hematopoietic progenitors and neural progenitors was less accessible than their corresponding mature cell

types in the single-cell assay for transposase accessible chromatin using sequencing (scATAC-seq) data of the mouse and human⁶¹⁻⁶³ (Fig. 5e, f).

***Xbp1* as a common regulator in multi-lineage progression.**

To dissect the mechanistic roles of the potential lineage-common regulators *Xbp1*, we used CRISPR/Cas9 to disrupt the *Xbp1* locus in mice (Fig. 6a, Extended Data Fig. 9a, b and Supplementary Table 69). As most *Xbp1*^{-/-} embryos died at E13.5, we applied scRNA-seq to analyze embryos at E12.5 from *Xbp1*^{+/-} heterozygous crosses before massive embryonic lethality⁶⁰ (Fig. 6b, c, Extended Data Fig. 9c and Supplementary Table 70). We found that increased cell groups after *Xbp1* disruption are all related with progenitor and immature cells (e.g., fetal mesenchymal progenitors, early primitive erythroid progenitor, muscle progenitors, radial glia, oligodendrocyte progenitors, and immature neurons) (Extended Data Fig. 9d). In addition, when compared to wild-type cells, *Xbp1*^{-/-} cells displayed higher entropy in a broad range of lineages which may lead to the eventual failure of cell type maturation (Fig. 6d and Extended Data Fig. 9e, f). Then we performed differential expression analysis and observed that a group of ribosomal protein genes (e.g., *Rps3a1* and *Rps7*) were specifically upregulated in *Xbp1*^{-/-} cells. Moreover, progenitor markers such as *Sox4*, *Id2*, *Son*, and the imprinted gene *H19* were enriched in *Xbp1*^{-/-} cells. The lineage-common regulators *Fosb* and *Jun* were downregulated in *Xbp1*^{-/-} cells (Fig. 6e, Supplementary Table 71). Thus, disruption of *Xbp1* caused mouse embryos to acquire a more stem-cell-like state.

To characterize the loss of function changes at protein levels, we performed liquid chromatography-mass spectrometry (LC-MS) proteomic analysis on both WT and KO embryos (Supplementary Table 72). *Xbp1*^{-/-} embryos exhibited higher expression level of pluripotency-related proteins such as *Lin28a*, *Lin28b*⁶⁴, *Pcgf6*⁶⁵, and *Jarid2*⁶⁶ and lower expression level of cell type-specific proteins such as *Sncal* in neural cells, *Clu* in stromal cells, *Afp* in hepatocytes, *C1qb* in macrophages, and *Blvrb* in erythroid cells (Fig. 6f and Extended Data Fig. 10a). In addition, canonical *Xbp1* targets related to the unfolded protein response (UPR)⁶⁷⁻⁶⁹ displayed no significant changes at the protein level (Extended Data Fig. 10b). Furthermore, *Xbp1* disruption in mouse embryonic stem cells did not alter stem cell culture and pluripotent gene expression, indicating that *Xbp1* transcriptional regulation of lineage decision is not downstream of UPR (Extended Data Fig. 10c, d, and Supplementary Table 73).

We applied VarID⁷⁰ to qualify lineage-determining factor changes in scRNA-seq datasets. These significantly variable TFs in *Xbp1*^{-/-} samples displayed an *Xbp1* binding motif in both scATAC-seq and chromatin immunoprecipitation sequencing (ChIP-seq) data (Fig. 6g and Supplementary Table 74). Our results indicate a direct role of *Xbp1* in lineage maturation via transcriptional regulation, during which *Xbp1* functions through a noncanonical mechanism that is independent of the UPR.

Discussion

Overall, our comprehensive atlas of mouse cell differentiation and maturation offers a powerful resource for investigating cell fate decisions. We characterize a general feature of decreased

entropy in most lineages along with development. Our analysis of gene network dynamics reveals both lineage-common and lineage-specific regulators that contribute to cell fate decisions. The proposal and definition of lineage-common regulators further refine a new functional classification of gene regulatory programs. We highlight that *Xbp1* is a critical and conserved transcriptional regulator of cell-type attractors in many lineages. A comprehensive and high-resolution analysis of the phenotype of *Xbp1* knockout was carried out in mouse embryonic development at the multi-omics level. However, the regulatory mechanisms of lineage-common regulators still require further research and functional validation on other systems such as in vitro differentiation, de-differentiation, and trans-differentiation.

Tissue development and maturation atlases may provide global views of cell fate decision process. Using our atlas data, we verified novel cell types with co-expression and ectopic expression patterns during mouse development. We hypothesize that early transitional cell types may serve as a pool of tissue-specific progenitors to widely support the normal progression of functional tissue formation. We also observe cell types that can't be annotated with existing knowledge. Lots of these unknown populations are seen in the neonatal mice. They need to be verified and explored in the future. By integrating developmental atlases across species, we illustrate common characteristics at varying evolutionary distances during development. We characterize a general feature of decreased entropy in most lineages along with development across species. Entropy is a concise, independent, and robust measurement for differentiation potential and we further associated it with ribosomal protein gene expression in evolutionarily distant species.

We identify *Xbp1* as a representative lineage-common regulator. We then used whole embryo scRNA-seq analysis to investigate the role of *Xbp1* across different lineages and cell types. The strategy provides a systematic insight into the gene function at organism level. Similar strategies can be applied to a series of knockout embryos for the dissection of a functional GRN during the development. It would also be interesting to compare quantitative gene function across different model systems.

In this study, we propose a systematic view of the cross-species state manifold landscape. Cells gradually progress from a stem/progenitor state towards specific cell fates with decreased entropy. During the process, there is an explosion of GRN diversity following lineage-specific differentiation. Importantly, we identify examples of common GRNs as conserved regulators of cell fate stabilization (Fig. 6h). In conclusion, the interpretation of our cell differentiation atlases leads to new understandings of cell fate decisions and the cellular state manifolds.

Acknowledgments

G.G. is a participant of the Human Cell Atlas Project. We thank Junqing Wu, Hangjun Wu, Jinfeng Liu, Dan Wang, Sheng Xiong, Ying Huang, and Pengfei Xu for support on the project. We thank G-BIO (Hangzhou), BGI (Shenzhen) and CNGB (Shenzhen) for supporting this project; Vazyme (Nanjing) for supplying the customized enzymes in the study; the Core Facilities of Zhejiang University School of Medicine for technical support; the Center of Cryo-Electron Microscopy at Zhejiang University for computational support; PTM Biolab (Hangzhou) for performing the liquid chromatography-mass spectrometry (LC-MS) proteomic analysis; and Gempharmatech (Nanjing) for producing the knockout mice. Funding: National

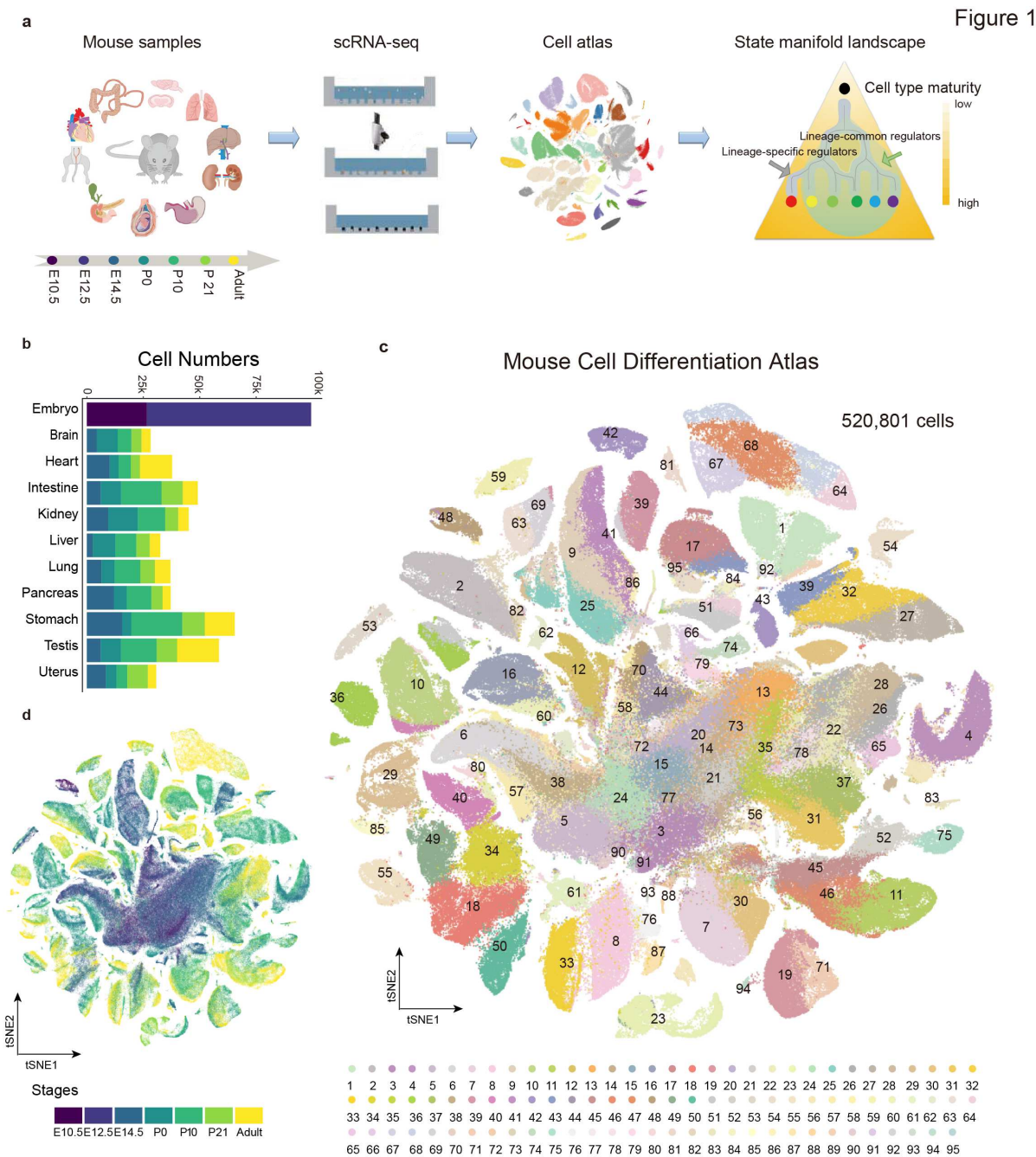
Natural Science Foundation of China (grants 31930028, 31871473, 31922049, 91842301, 32000461, 32001068, T2121004), National Key Research and Development Program (grants 2018YFA0107804, 2018YFA0107801, 2018YFA0800503), Fundamental Research Funds for the Central Universities (G.G.).

Author Contributions

The project was conceived by G.G. Experiments were performed by X.H., H.C., X.F., Z.Z., R.W. and L.F. Single-cell data processing, clustering analyses, gene regulated analyses and cell type annotation were performed by L.F., L.M., W.E., H.S., J.W., X.W., X.Y., and Y.M. The paper was written by G.G., L.F., H.C., L.M., W.E., and X.F. Sequencing experiments were performed by M.J., D.J., and T.Z. Preserving and making available data, code and materials upon publication were done by L.F., L.M., and W.E. Funding was acquired by G.G., X.H., H.C., J.W.

Competing Interests

Authors declare that they have no competing interests.



335

336 **Figure 1. Single-cell transcriptional atlas of mouse differentiation.**

337 **a**, Overview of the experimental and bioinformatics analysis workflow. **b**, A total of 10 organs
338 were analyzed at 7 different time points. The bar plot shows the number of sequenced cells per
339 organ prepared by Microwell-seq. **c**, *t*-SNE visualization of 520,801 single cells from the
340 Mouse Cell Differentiation Atlas, colored by cluster identity. **d**, *t*-SNE visualization of 520,801
341 single cells from different developmental stages of mice, colored by stages. Fig. 1b, d share the
342 same color legend of stages.

343

Figure 2

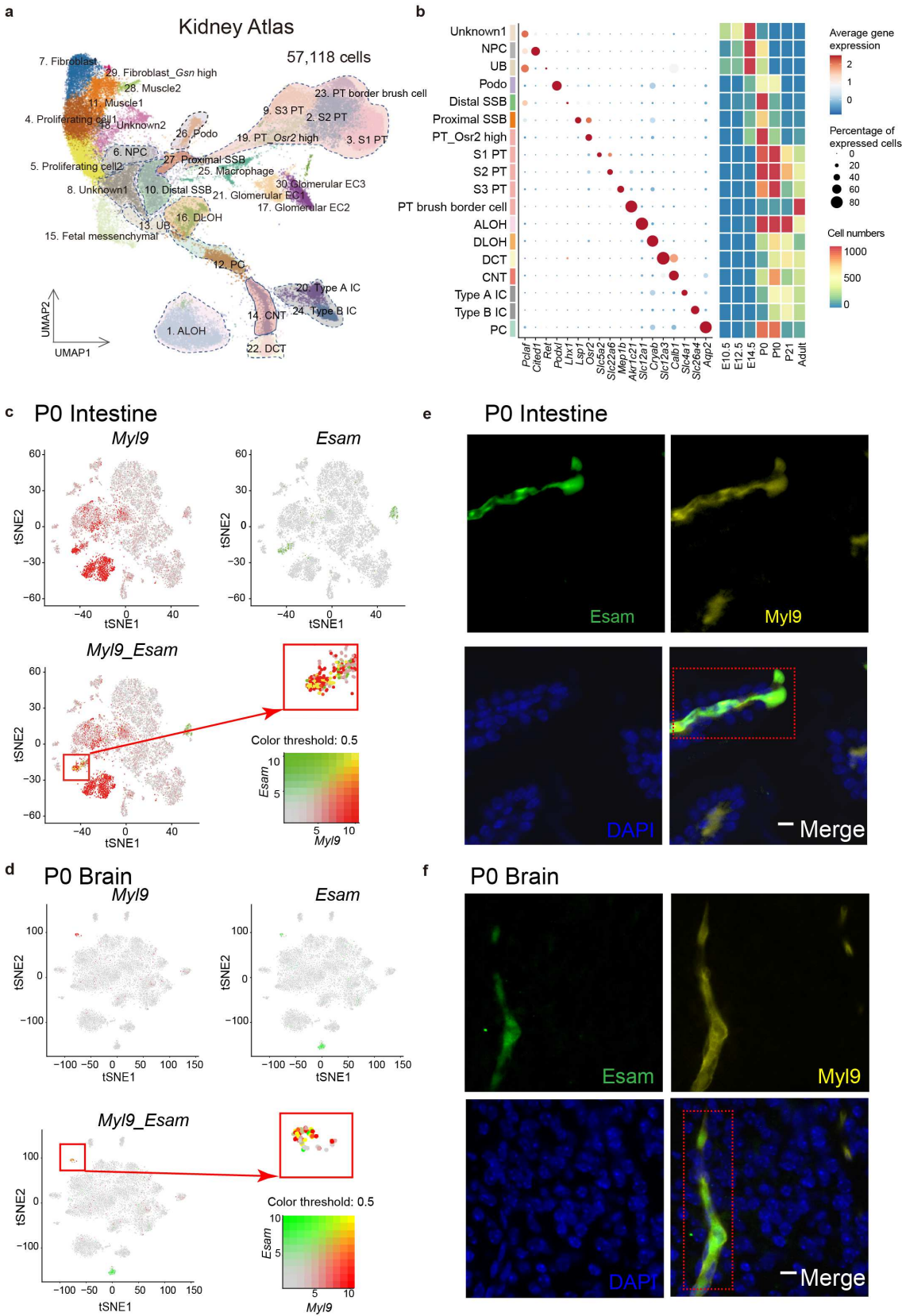


Figure 2. Cellular heterogeneity in mouse tissues.

a, UMAP visualization of 57,118 single cells in the kidneys at 7 different time points, colored by cluster identity. **b**, Dot plot visualization of representative markers expression in each cell

type in the kidney single-cell data. The size of the dot encodes the percentage of cells within the cell type, and the color encodes the average expression level. Heatmap showing the cell number of corresponding cell types at each time point. **c, d**, Feature plots in the t-SNE map of P0 intestine (**c**, $n = 9,265$ cells) and P0 brain (**d**, $n = 9,101$ cells). Cells are colored according to the expression of the indicated marker genes or two genes. The red boxes magnify the co-expressed cell types in the tissues. **e, f**, Immunofluorescence assay for the cells that co-expressed makers of myocytes (MyI9) and endothelial cells (Esam) in both intestine (**e**) and brain (**f**) at the P0 stage. The blue marks the cell nucleus using DAPI. The red boxes indicate the co-expressed locations. The experiment was replicated three times with similar results. Scale bar, 20 μm .

Figure 3

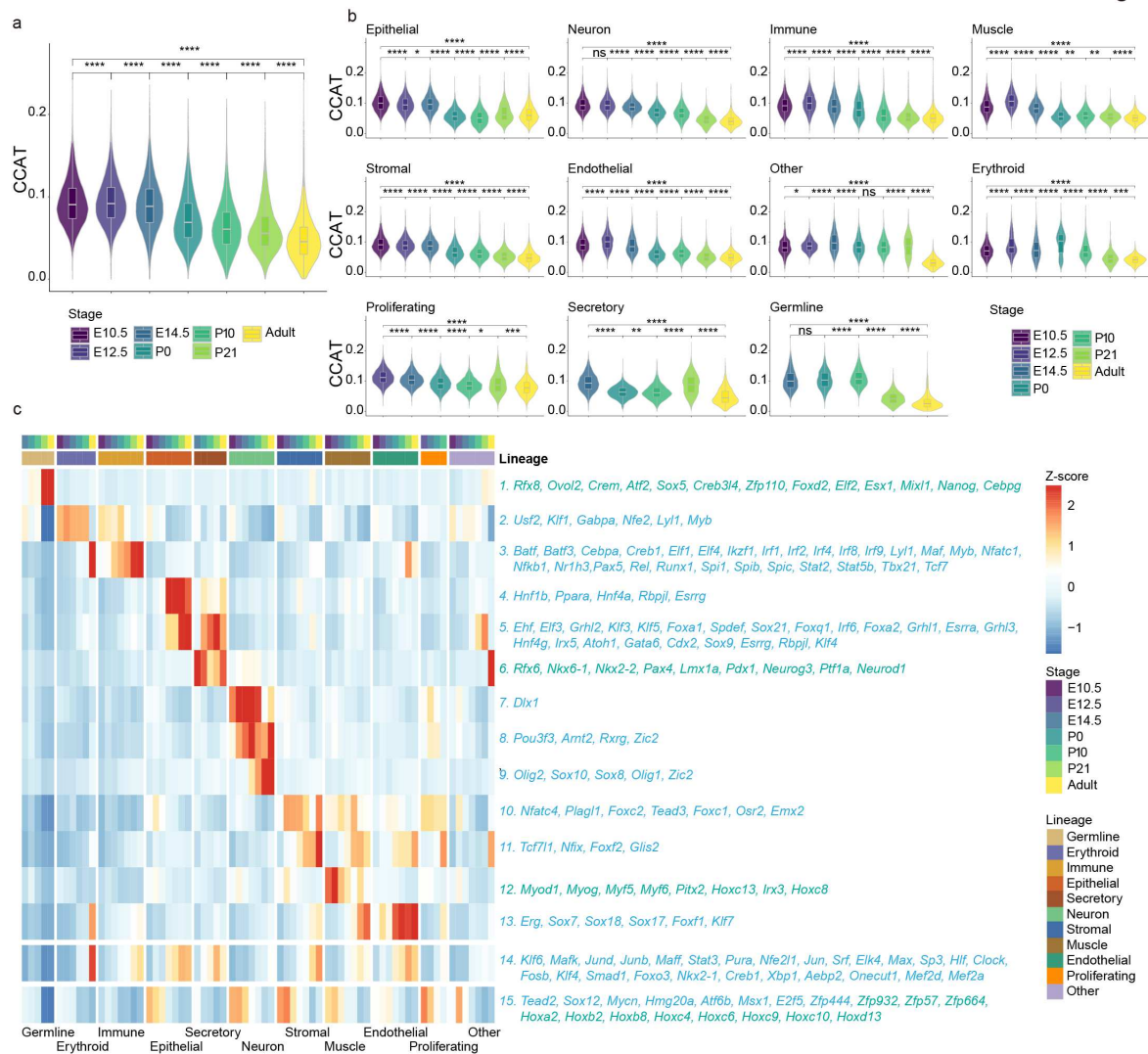


Figure 3. Analysis of regulatory programs in MCDA.

a, Entropy measurement of MCDA using the CCAT method in different development stages ($n = 520,801$ cells). P-values are from a two-sided Wilcoxon rank sum test comparing entropies of two different development stages (ns: no significance, p-value > 0.05 , * p-value ≤ 0.05 , ** p-value ≤ 0.01 , *** p-value ≤ 0.001 , **** p-value ≤ 0.0001). The exact p values were

displayed in the Source Data. Box plots: center line, median; boxes, first and third quartiles of the distribution; whiskers, highest and lowest data points within $1.5 \times \text{IQR}$. The same statistical analysis was performed for Fig. 3a, b. **b**, Entropy measurement of each lineage in MCDA using the CCAT method in different development stages (epithelial: n = 116,436 cells, neuron: n = 41,342 cells, immune: n = 75,433 cells, muscle n = 17,909 cells, stromal: n = 106,955 cells, endothelial: n = 23,243 cells, other: n = 30,575 cells, erythroid: n = 41,683 cells, proliferating: n = 16,567 cells, secretory: n= 15,161 cells, germline: n = 35,497 cells). **c**, Heatmap of aggregated module activities of TFs clustered by fuzzy c-means showing variations by stage and lineage from SCENIC. Representative TFs of each lineage in MCDA are listed. Blue marks the TFs in collection A (high confidence) and green (medium confidence) marks the TFs in collection B.

376

Figure 4

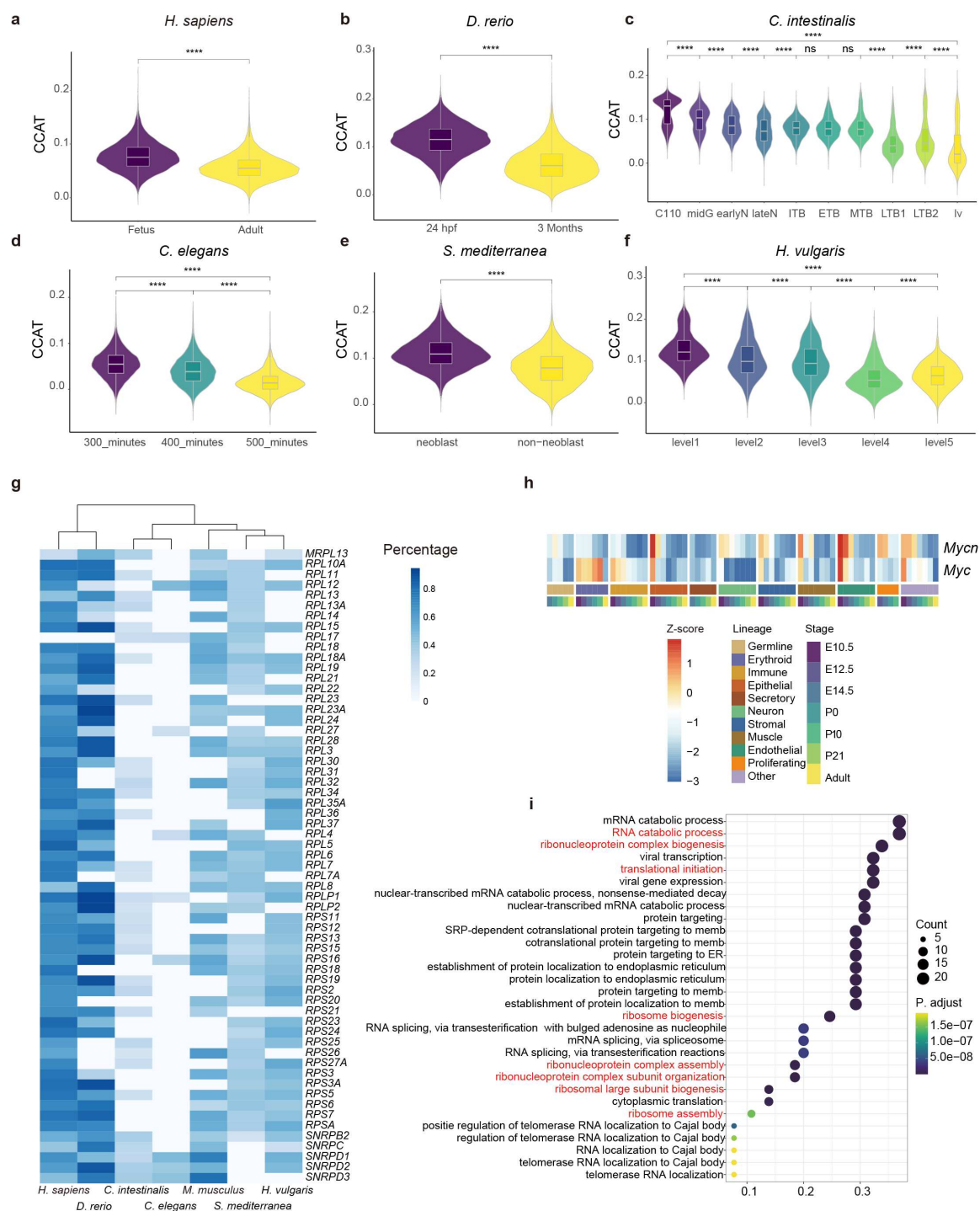
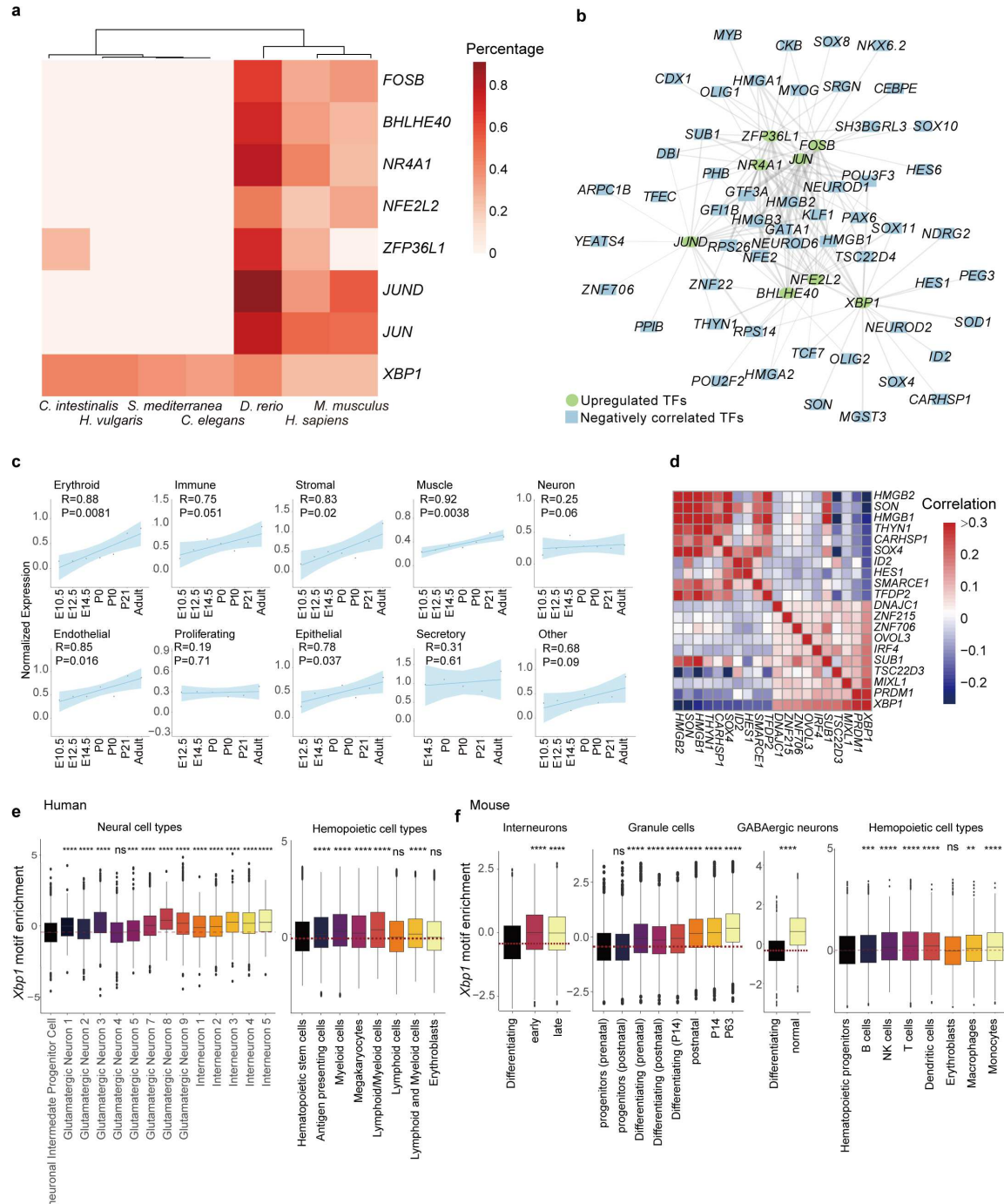


Figure 4. Global characteristics of the cell differentiation across species.

a-f, Entropy measurement of cells in *H. sapiens* (**a**, $n = 85,181$ cells), *D. rerio* (**b**, $n = 76,838$ cells), *C. intestinalis* (**c**, $n = 90,579$ cells), *C. elegans* (**d**, $n = 61,810$ cells), *S. mediterranea* (**e**, $n = 21,612$ cells) and *H. vulgaris* (**f**, $n = 25,052$ cells) using the CCAT methods. The color represents the stage. P-values are from a two-sided Wilcoxon rank sum test comparing entropies of two different development stages (ns: no significance, p-value > 0.05, * p-value <= 0.05, ** p-value <= 0.01, *** p-value <= 0.001, **** p-value <= 0.0001). The exact p values were displayed in the Source Data. Box plots: center line, median; boxes, first and third

quartiles of the distribution; whiskers, highest and lowest data points within $1.5 \times \text{IQR}$. **g**, Heatmap showing the cell type frequencies of commonly downregulated ribosomal protein genes, mitochondrial ribosomal proteins, and small nuclear ribonucleoprotein genes in at least four species. More genes are included in the Supplementary Table 64. **h**, Heatmap showing the activity scores of *Mycn* and *Myc* in different stages and lineages in MCDA. **i**, Gene enrichment analysis of the driving genes in CCAT method. The top 30 enriched biological processes were displayed. Red marks the GO terms related to the ribosome biogenesis.

Figure 5

**Figure 5. Inference of gene regulation during cell fate decisions across species.**

a, Heatmap showing the cell type frequencies of commonly upregulated TFs in 7 species. **b**, Regulator network showing the top 20 most negatively relevant TFs in the HCL for the commonly upregulated TFs in Fig. 5a, (Pearson correlation p-value ≤ 0.05). **c**, Scatter plot showing aggregated *Xbp1* expression patterns in MCDA per lineage. Lines were estimated through linear regression and the 95% confidence interval is shown in blue with the mean value in grey points. **d**, Heatmap showing the top 10 TFs most correlated with *XBP1* in the HCL. **e-f**, Box plot showing the z scores of *Xbp1* motif enrichment in neural cell types and hemopoietic cell types in the human (**e**) and mouse (**f**) in scATAC-seq data (Human neural cell types: n =

22,075 cells, Human hemopoietic cell types: n = 16,133 cells, mouse interneurons: n = 5,134 cells, mouse granule cells: n = 25,155 cells, mouse GABAergic neurons: n = 2,041 cells, mouse hemopoietic cell types: n = 24,125 cells). P-values are from a two-sided Wilcoxon rank sum test comparing the *Xbp1* enrichment score between the progenitor cell types (the first box) and other cell types (ns: no significance, p-value > 0.05, * p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001, **** p-value ≤ 0.0001). The exact p values were displayed in the Source Data. Box plots: center line, median; boxes, first and third quartiles of the distribution; whiskers, highest and lowest data points within $1.5 \times \text{IQR}$.

Figure 6

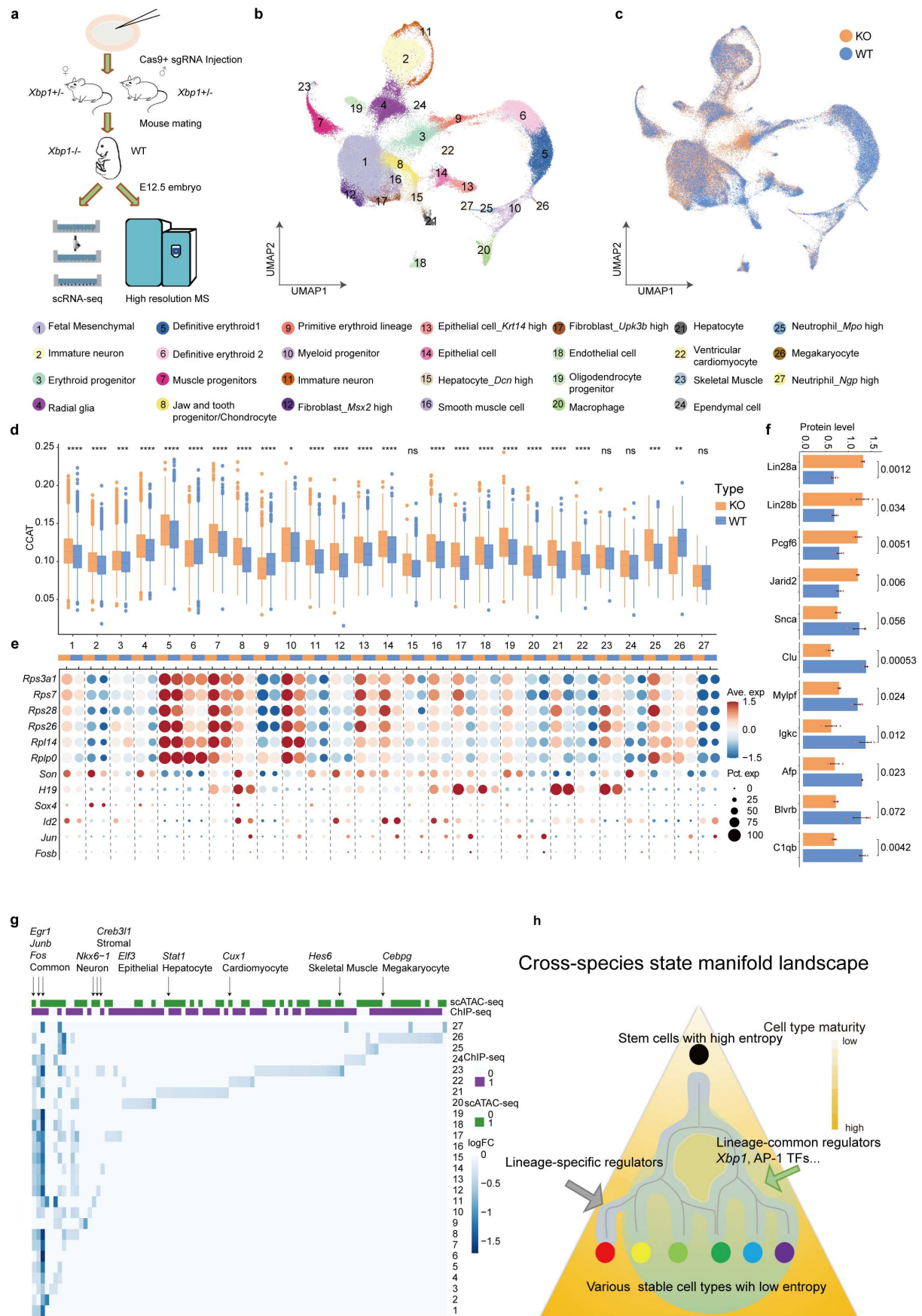


Figure 6. scRNA-seq and high-resolution MS revealed gene and protein changes in *Xbp1*^{-/-} embryos.

a, Overview of the CRISPR/Cas9 experiment. *Xbp1*^{-/-} and wild-type embryos at E12.5 were prepared and processed by Microwell-seq and liquid chromatography-mass spectrometry (LC-MS). **b, c**, UMAP visualization of 93,246 single cells from *Xbp1*^{-/-} and wild-type embryos at E12.5, colored by cluster identity (**b**) and genotype (**c**) (KO: n = 49, 498, WT: n = 43,748). **d**, Entropy measurement of each cluster in *Xbp1*^{-/-} and wild-type embryos using the CCAT method (n = 93,246 cells). The color represents the genotype. P-values are from a two-sided Wilcoxon rank sum test comparing entropies of two groups with different genotypes from the same cluster (ns: no significance, p-value > 0.05, * p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001, **** p-value ≤ 0.0001). The exact p values were displayed in the Source Data. Box plots: center line, median; boxes, first and third quartiles of the distribution; whiskers, highest and lowest data points within 1.5 × IQR. **e**, Dot plot showing representative differentially expressed genes (ribosomal protein genes, progenitor marker genes, and lineage-common regulators) of each cluster in KO and WT cells. **f**, Bar plot showing representative protein expression levels (for pluripotency-related proteins and cell type-specific proteins) between *Xbp1*^{-/-} and wild-type mice (KO: n = 3, WT: n = 3, mean ± s.d.). A two-sided t-test was performed to determine the statistical significance. The illustrative genes were manually selected from the full heatmap, which is shown in Extended Data Figure. 10a. **g**, Heatmap showing significantly variable TFs in *Xbp1*^{-/-} samples. Green and purple indicate the chromatin accessibility of the *Xbp1* binding motif as determined by scATAC-seq and ChIP-seq, respectively. The *Xbp1* binding motif of the mouse was from CisBP database. Representative TFs are marked and were manually selected from Supplementary Table 74. **h**, Schematic of cross-species state manifold landscape.

References

1. Mathis, L. & Nicolas, J.-F. Cellular patterning of the vertebrate embryo. *TRENDS in Genetics* **18**, 627–635 (2002).
2. Heinäniemi, M. *et al.* Gene-pair expression signatures reveal lineage control. *Nat Methods* **10**, 577–583 (2013).
3. Moris, N., Pina, C. & Arias, A. M. Transition states and cell fate decisions in epigenetic landscapes. *Nature Reviews Genetics* **17**, 693–703 (2016).
4. Waddington, C. H. *The strategy of the genes*. (Routledge, 2014).
5. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat Rev Genet* **21**, 410–427 (2020).

448 6. Teschendorff, A. E. & Feinberg, A. P. Statistical mechanics meets single-cell biology. *Nat Rev Genet* **22**,
449 459–476 (2021).

450 7. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell Fates as High-Dimensional Attractor States of a
451 Complex Gene Regulatory Network. *Phys. Rev. Lett.* **94**, 128701 (2005).

452 8. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **132**, 631–644
453 (2008).

454 9. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts
455 to myoblasts. *Cell* **51**, 987–1000 (1987).

456 10. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**,
457 917–929 (2005).

458 11. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).

459 12. Plass, M. *et al.* Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics.
460 *Science* **360**, (2018).

461 13. Siebert, S. *et al.* Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science*
462 **365**, (2019).

463 14. Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).

464 15. Packer, J. S. *et al.* A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution.
465 *Science* **365**, (2019).

466 16. Cao, C. *et al.* Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354
467 (2019).

468 17. Mittnenzweig, M. *et al.* A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**,
469 2825–2842. e22 (2021).

18. Qiu, C. *et al.* Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat Genet* **54**, 328–341 (2022).
19. Ferre, P., Decaux, J.-F., Issad, T. & Girard, J. Changes in energy metabolism during the suckling and weaning period in the newborn. *Reproduction Nutrition Développement* **26**, 619–631 (1986).
20. Stewart, B. J. *et al.* Spatiotemporal immune zonation of the human kidney. *Science* **365**, 1461–1466 (2019).
21. Chen, L. *et al.* Renal-Tubule Epithelial Cell Nomenclature for Single-Cell RNA-Sequencing Studies. 8 (2019).
22. Zheng, B. *et al.* Prospective identification of myogenic endothelial cells in human skeletal muscle. *Nat Biotechnol* **25**, 1025–1034 (2007).
23. Chaudhari, P., Tian, L., Deshmukh, A. & Jang, Y.-Y. Expression kinetics of hepatic progenitor markers in cellular models of human liver development recapitulating hepatocyte and biliary cell fate commitment. *Experimental Biology and Medicine* **241**, 1653–1662 (2016).
24. Willnow, D. *et al.* Quantitative lineage analysis identifies a hepato-pancreato-biliary progenitor niche. *Nature* **597**, 87–91 (2021).
25. Banerji, C. R. S. *et al.* Cellular network entropy as the energy potential in Waddington’s differentiation landscape. *Sci Rep* **3**, 3039 (2013).
26. Grün, D. *et al.* De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19**, 266–277 (2016).
27. Guo, M. SLICE: determining cell differentiation and lineage based on single cell entropy. *Nucleic Acids Research* **45**, 14 (2017).
28. Teschendorff, A. E. & Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell’s transcriptome. *Nature Communications* **8**, 15599 (2017).

492 29. Teschendorff, A. E., Maity, A. K., Hu, X., Weiyan, C. & Lechner, M. Ultra-fast scalable estimation of
493 single-cell differentiation potency from scRNA-Seq data. *Bioinformatics* **37**, 1528–1534 (2021).

494 30. Kauffman, S. Homeostasis and differentiation in random genetic control networks. *Nature* **224**, 177–178
495 (1969).

496 31. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251
497 (2013).

498 32. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**,
499 1083–1086 (2017).

500 33. Holland, C. H. *et al.* Robustness and applicability of transcription factor and pathway analysis tools on
501 single-cell RNA-seq data. *Genome Biol* **21**, 36 (2020).

502 34. Parker, H. J. Mammalian Embryo: Hox Genes. *eLS* 1–15 (2020).

503 35. Cassandri, M. *et al.* Zinc-finger proteins in health and disease. *Cell death discovery* **3**, 1–12 (2017).

504 36. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *TRENDS in Genetics* **29**, 569–574
505 (2013).

506 37. Jochum, W., Passegué, E. & Wagner, E. F. AP-1 in mouse development and tumorigenesis. *Oncogene* **20**,
507 2401–2412 (2001).

508 38. Velazquez, F. N., Caputto, B. L. & Boussin, F. D. c-Fos importance for brain development. *Aging (Albany*
509 *NY)* **7**, 1028 (2015).

510 39. Liu, J. *et al.* The oncogene c-Jun impedes somatic cell reprogramming. *Nature cell biology* **17**, 856–867
511 (2015).

512 40. Madrigal, P. & Alasoo, K. AP-1 takes centre stage in enhancer chromatin dynamics. *Trends in cell biology*
513 **28**, 509–511 (2018).

514 41. Hinman, V. & Cary, G. The evolution of gene regulation. *Elife* **6**, e27291 (2017).

42. Li, J. *et al.* Inferring predictive genetic models and regulatory elements by deep learning of cross-species single-cell gene expression landscapes. 39 (2021). (Submitted to Nature Genetics as NG-LE58036R2 and also in the process of preprinting in Research Square.)
43. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology* **20**, 1–9 (2019).
44. Briggs, J. A. *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, (2018).
45. Chakraborty, C. & Agoramoorthy, G. Stem cells in the light of evolution. *The Indian journal of medical research* **135**, 813 (2012).
46. Saeed, R. & Deane, C. M. Protein protein interactions, evolutionary rate, abundance and age. *BMC bioinformatics* **7**, 1–13 (2006).
47. Boon, K. *et al.* N- myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *The EMBO journal* **20**, 1383–1393 (2001).
48. Van Riggelen, J., Yetil, A. & Felsher, D. W. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nature Reviews Cancer* **10**, 301–309 (2010).
49. Shi, J., Teschendorff, A. E., Chen, W., Chen, L. & Li, T. Quantifying Waddington’s epigenetic landscape: a comparison of single-cell potency measures. *Briefings in bioinformatics* **21**, 248–261 (2020).
50. Farley-Barnes, K. I. *et al.* Diverse regulators of human ribosome biogenesis discovered by changes in nucleolar number. *Cell reports* **22**, 1923–1934 (2018).
51. Macaulay, I. C. *et al.* Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep* **14**, 966–77 (2016).
52. Athanasiadis, E. I. *et al.* Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. *Nat Commun* **8**, 2045 (2017).

53. Krausgruber, T. *et al.* Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296–302 (2020).
54. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols* **15**, 2247–2276 (2020).
55. Wang, J. *et al.* Tracing cell-type evolution by cross-species comparison of cell atlases. *Cell Reports* **34**, 108803 (2021).
56. Lee, A.-H., Chu, G. C., Iwakoshi, N. N. & Glimcher, L. H. XBP-1 is required for biogenesis of cellular secretory machinery of exocrine glands. *EMBO J* **24**, 4368–4380 (2005).
57. Todd, D. J. *et al.* XBP1 governs late events in plasma cell differentiation and is not required for antigen-specific memory B cell development. *Journal of Experimental Medicine* **206**, 2151–2159 (2009).
58. Pramanik, J. *et al.* Genome-wide analyses reveal the IRE1a-XBP1 pathway promotes T helper cell differentiation by resolving secretory stress and accelerating proliferation. *Genome Med* **10**, 76 (2018).
59. Masaki, T., Yoshida, M. & Noguchi, S. Targeted Disruption of CRE-Binding Factor TREB5 Gene Leads to Cellular Necrosis in Cardiac Myocytes at the Embryonic Stage. *Biochemical and Biophysical Research Communications* **261**, 350–356 (1999).
60. Reimold, A. M. *et al.* An essential role in liver development for transcription factor XBP-1. *Genes & development* **14**, 152–157 (2000).
61. Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309–1324.e18 (2018).
62. Di Bella, D. J. *et al.* Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature* **595**, 554–559 (2021).
63. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).

64. Viswanathan, S. R. & Daley, G. Q. Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445–449 (2010).
65. Yang, C.-S., Chang, K.-Y., Dang, J. & Rana, T. M. Polycomb group protein Pcgf6 acts as a master regulator to maintain embryonic stem cell identity. *Scientific reports* **6**, 1–12 (2016).
66. Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306–310 (2010).
67. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA Is Induced by ATF6 and Spliced by IRE1 in Response to ER Stress to Produce a Highly Active Transcription Factor. *Cell* **107**, 881–891 (2001).
68. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882.e21 (2016).
69. Acosta-Alvear, D. *et al.* XBP1 Controls Diverse Cell Type- and Condition-Specific Transcriptional Regulatory Networks. *Molecular Cell* **27**, 53–66 (2007).
70. Grün, D. Revealing dynamics of gene expression variability in cell state space. *Nat Methods* **17**, 45–49 (2020).

Methods

Mouse experiments to supplement the MCDA database

Wild-type C57BL/6J mice were ordered from Shanghai Laboratory Animal Center (SLAC, Shanghai). All mice were housed at Zhejiang University Laboratory Animal Center in a specific pathogen-free (SPF) facility with individually ventilated cages. The room had a controlled temperature (20-22 °C), humidity (30-70%), and light program (12 h light-dark cycle). The mice were provided ad libitum access to a regular rodent chow diet.

To obtain embryonic samples (E10.5 embryos, E12.5 embryos), C57BL/6 mice were mated. Noon on the day the vaginal plug was visible was considered to be E0.5. Sex was not determined before tissue pooling due to the small sample sizes for E10.5, E12.5, and P0 samples (except for the gonads). Embryos were collected from at least three independent litters per development stage. For P10 and P21 samples, testes were collected from male mice, and all the other tissues were collected from female mice.

All experiments performed in this study were approved by the Animal Ethics Committee of Zhejiang University. All experiments conformed to the relevant regulatory standards at Zhejiang University Laboratory Animal Center.

Generation of Xbp1 knockout mESC and mouse models

SgRNAs targeting exon 2 of *Xbp1* were designed using the Zhang laboratory CRISPR design website tool (<http://crispr.mit.edu/>). Oligonucleotides were synthesized and then cloned into an epiCRISPR/Cas9 vector⁷¹. The vector was extracted using an EndoFree Mini Plasmid Kit II (4992422, Tiangen Biotech) following the manual. Approximately 4×10^5 E14 mouse embryonic stem cells were transfected with 2 μ g of the vector with Lipofectamine 3000 (L3000001, Life Technologies) based on an online protocol. At days 2-10, cells were selected with puromycin (0.5-1.0 μ g/ml). Then, single cells were reseeded in a 6-well plate and cultured in mESC mediums for 7-10 days. Individual colonies were picked and genotyped. The gRNA target sites and oligonucleotides used in this study can be found in Supplementary Table 73.

Xbp1 knockout C57BL/6J mice were generated by Nanjing Gempharmatech. Mice were genotyped by PCR using genomic tail DNA. To obtain live knockout embryos at E12.5 for scRNA-seq, we used a ScientificTM PhireTM Animal Tissue Direct PCR Kit (F140WH, Thermo) to genotype embryos quickly. All primers used for knockout and genotyping are listed in Supplementary Table 69.

Immunofluorescent staining

Fresh mouse tissues were frozen in disposable molds containing OCT. Frozen sections were cut at 10 μ m in CryoStar NX50 (Thermo), mounted on microscope slides and stored at -80 °C. Before staining, the sections were thawed for 20 min and 4% formaldehyde in PBS was added to cover the sections. Tissues were fixed for 15 min at room temperature. After fixation, sections were washed three times with PBS. Cells were permeabilized by covering the sections with 0.1% Triton X-100 in PBS for 10 min. Then, the sections were washed three times with PBS and blocked with 3% BSA in PBS for 1 h at room temperature. Primary antibodies (anti-ESAM (1:50, MA5-24072; Thermo), anti-Myl9 (1:400, ab187152; Abcam), anti-Scgb1a1 (1:50, MAB4218-SP; R&D), anti-tff2 (1:200, 13681-1-AP; ProteinTech), anti-AFP (1:200, AF5134; Affinity)) diluted in blocking solution were added to cover the sections. The slides were placed in a wet box and incubated overnight at 4 °C. Relevant AlexaFluor488/594-conjugated secondary antibodies (1:1,000, A-21208, A-21206 and A-11037, Thermo) were used for labeling. The slides were then washed three times with blocking solution and stained with DAPI. Glass coverslips were then attached to the slides using mounting mediums. Immunofluorescence images were obtained using Olympus VS200.

Western blot

The mouse embryos were solubilized in RIPA lysis buffer (20mg/200ul, P0013D, Beyotime). The mixture was lysed using a homogenizer for 5 min on ice. Tissue lysates were then cleared by centrifugation at 14,000g for 10 min at 4 °C. Equal amounts of total protein were used for experimental and control. Samples were fractionated by SDS-PAGE and transferred to polyvinylidene-fluoride (PVDF) membrane. After blocking with 5% milk in TBST for 1 h at room temperature, the membranes were probed with the corresponding primary and secondary antibodies. Primary antibodies (anti-Xbp1 (1:1000, ab37152; Abcam), anti- β -tubulin (1:3000,

EM0103; HUABIO)) and secondary antibodies (anti-mouse IgG (1:5000, HS201-01; TransGen Biotech), anti-rabbit IgG (1:5000, GAR007; Multi Science)) diluted in TBST were used.

Cell preparation

Mouse tissues were minced into pieces of ~1 mm on ice using scissors. The tissue pieces were transferred to a 15 ml centrifuge tube, rinsed twice with cold DPBS, and suspended in 5 ml of a solution containing dissociation enzymes. The samples were treated with various enzymes for different amounts of time (Supplementary Table 3). During dissociation, the tissue pieces were pipetted up and down gently several times until no tissue fragments were visible. The dissociated cells were centrifuged at 300×g for 5 min at 4 °C and then resuspended in 3 ml of cold DPBS. After passage through a 40-μm strainer (Biologix), the cells were washed twice, centrifuged at 300×g for 5 min at 4 °C, and resuspended at a density of 1×10⁵ cells/ml in cold DPBS containing 2 mM EDTA.

Single-cell RNA-sequencing

Single-cell cDNA libraries were prepared using the Microwell-seq. Briefly, cells were loaded on the microwell plate and extra cells were washed away using ice-cold PBS gently. Then bead suspension (sequences listed in Supplementary Table 2) was loaded on the plate and extra beads were washed away on a magnet. The plate was covered using cold lysis buffer (0.1 M Tris-HCl pH 7.5, 0.5 M LiCl, 1% SDS, 10 mM EDTA, and 5 mM dithiothreitol) and incubated on ice for 12 min. Then, beads were collected and washed using 6× SSC and 50 mM Tris-HCl pH 8.0. After washing, beads were resuspended in RT mix and incubated at 42°C for 90 min. After RT, beads were washed in TE-TW and 10 mM Tris-HCl pH 8.0. Beads were resuspended in exonuclease I mix and incubated at 37°C for 30 min. Then, beads were washed in TE-SDS, TE-TW, and 10 mM Tris-HCl pH 8.0. Beads were resuspended in PCR mix with TSO primer to amplify the cDNA. After PCR, beads were removed and cDNA products were purified using 0.8 X VAHTS DNA Clean Beads (N411-01, Vazyme). A more detailed version of the Microwell-seq protocol is available in Han et al¹⁴. Then, the purified cDNA libraries were fragmented using a customized transposase that carries two identical insertion sequences. The customized transposase was included in the TruePrep Homo-N7 DNA Library Prep Kit for Illumina (TD513, Vazyme) or TruePrep Homo-N7 DNA Library Prep Kit for MGI (L-N7E461L0, Vazyme). The fragmentation reaction was performed according to the instructions provided by the manufacturer. We used customized P5 primer (listed in Supplementary Table 2) and VAHTS RNA Adapters set 3-set 6 for Illumina (N809/N810/N811/N812, Vazyme) or our MGI P7 primers (N8XX, listed in Supplementary Table 2) to specifically amplify fragments that contain the 3' ends of transcripts. Other fragments will form self-loops, impeding their binding to PCR primers. The PCR program was as follows: 72°C for 3 min; 98°C for 1 min; five cycles of 98°C for 15 s, 60°C for 30 s, and 72°C for 3 min; 72°C for 5 min; and a 4°C hold. The PCR product was purified using 0.9X VAHTS DNA Clean beads (N411-01, Vazyme). Then, a 25 μL PCR mix (1×HiFi HotStart Readymix and 0.2 μM 2100 primer) was added to each sample. The PCR program was as follows: 95°C for 3 min; five cycles of 98°C for 20 s, 60°C for 15 s, and 72°C for 15 s; 72°C for 3 min; and a 4°C hold. To eliminate primer dimers and large fragments, 0.55-0.15X VAHTS DNA Clean beads were then used to purify the cDNA library. The size distribution of the products was analyzed on an Agilent 2100 bioanalyzer, and

a peak in the 400 to 700 bp range was observed. Finally, the samples were subjected to sequencing on the Illumina HiSeq (data for MDCA) or MGI DNBSEQ-T7 (data for *Xbp1* knockout experiment). For MGI sequencing, we applied the protocol provided by VAHTS Circularization Kit for MGI (NM201-01, Vazyme) to obtain single-stranded circular cDNA available for DNB (DNA Nanoball) generation. We also replaced the official R1 sequencing primers with our custom R1 sequencing primers A&B (listed in Supplementary Table 2) to ensure the completion of the sequencing.

Processing of Microwell-seq data

Microwell-seq data sets were processed as described¹¹. Reads were aligned to the *Mus_musculus*. GRCm38.88 genome using STAR⁷² (v2.5.2a). The DGE data matrices were obtained using the Drop-seq core computational protocol (available at website <http://mccarrolllab.org/dropseq/>) with the default parameters. For quality control, we filtered out cells with detection of fewer than 500 transcripts. Cells with a high proportion of transcript counts (>20%) derived from mitochondria-encoded genes were also excluded. Cells were also corrected for RNA contamination and background-removed DGE data were constructed¹⁴. The SCANPY⁷³ (v1.6.0) python package and Seurat⁷⁴ (v3.2.2) R package were used to load the cell-gene count matrix and perform downstream analysis.

Clustering of the single-cell data matrix

For clustering of the complete mouse tissue dataset (520,801 cells), qualified cells were processed using SCANPY in a Python environment. Background-removed DGE data for cells analyzed in each tissue and genes expressed in at least 20 cells were used as inputs¹⁴. Then, DGE data were $\ln(\text{CPM}/100 + 1)$ transformed. We selected approximately 3,000 highly variable genes according to their average expression and dispersion. We then regressed out UMI and gene numbers and scaled each gene to unit variance and the values beyond a standard deviation of 10 were clipped. For the mouse tissue dataset, we chose PCs for PCA according to elbow plots, and 50 PCs were used to create a neighborhood graph for the cells. We then used Leiden clustering to cluster with resolution = 8 and k = 25. Marker genes were calculated by the Wilcoxon rank-sum test (two-sided) and p-value adjustment was performed using benjamini-hochberg correction. For visualization, t-distributed stochastic neighbor embedding (t-SNE) was used.

For kidney data, bbknn⁷⁵ (v1.4.0) was performed by using ridge regression to remove batch effects. For clustering of single tissues, the Seurat pipeline was used with the default parameters for fewer cells. Cell type and lineage information of each cell type were manually annotated according to the marker genes reported in the previous paper¹¹. A hierarchical tree of the MCDA was computed using the correlations of average gene expression of 95 clusters with highly variable genes.

Estimation of the variance of the MCDA

To estimate the variance in the data depending on age, tissue, or sex, we first aggregated the gene expression for each tissue at multiple time points. Using the above metadata as input, we performed principal variance component analysis (PVCA) using R Package pvca (v1.26.0, <https://www.bioconductor.org/packages/release/bioc/html/pvca.html/>) with the default parameters. It leverages the strengths of two popular data analysis methods: principal

components analysis and variance components analysis, and integrates them into a novel algorithm. And it uses the eigenvalues associated with their corresponding eigenvectors as weights, to quantify the magnitude of each source of variability. All factors as well as their interaction terms are treated as random effects in the mixed model for variance component estimation. It fits a linear mix-effects model to data. Items like “tissue and gender” are variances explained by interactions of two factors instead of the union of two factors.

Inference of the TFs for MCDA

As a proof of principle, we applied experimental verified tissue-specific TFs from literature⁷⁶ as the golden standard. We included both tissue-restricted TFs and not uniformed expressed TFs in different tissues as tissue-specific TFs. For datasets used, we selected high-quality cells with more than 800 gene numbers as single-cell datasets, and also aggregated every 20 single cells in each cell type to produce pseudo cells to enrich our choices of input datasets. We compared SCENIC³² (v0.10.0) and VIPER-DOROTHEA³³ (viperv1.28.0 and dorothea v1.6.0) for inferring specific TFs in the tissues. The DOROTHEA database provided TFs from different types of evidence with a different confidence. We used ABCDE (1,113 TFs) categories of DOROTHEA TFs in our comparison. Regulon specificity scores (RSS)⁵⁴ were calculated to represent TF specificity in the tissue for both VIPER-DOROTHEA and SCENIC. Then we employed the youden index (sensitivity + specificity - 1) to find the best performance of VIPER-DOROTHEA and SCENIC in classifying tissue-specific TFs both in sensitivity and specificity. These TFs were compared with the golden standard lists in four aspects: sensitivity, specificity, FPR (false positive rate), and PRAUC (Area Under the Precision-Recall Curve).

To define regulatory programs in MCDA, SCENIC and VIPER-DOROTHEA were applied firstly to infer the gene regulatory network with default parameters using high-quality single cells with more than 800 genes. For VIPER-DOROTHEA, ABCDE (1113 TFs) categories of DOROTHEA TFs were used. Secondly, z-scaled RSS scores for VIPER-DOROTHEA and SCENIC in each stage-lineage were calculated as TF-by-lineage matrix. Then, fuzzy c-means clustering was performed on the TF-by-lineage matrix calculated by SCENIC and VIPER-DOROTHEA, resulting in a TF-by-module “membership matrix” and a lineage-by-module “centers matrix”. The “centers matrix” with 15 modules was used to generate the heatmap. We defined a threshold membership score (threshold = 0.2) in which TFs were assigned to a module. With the fuzzy c-means heatmap, we identified which modules/TFs were lineage-specific and which were lineage-sharing. We assigned TFs into specific lineages according to the aggregated patterns of modules manually and the resulting TFs were classified into three collections with high to low confidence: collection A consisted of TFs from both methods, collection B were TFs only from SCENIC, and collection C were TFs only from VIPER-DOROTHEA (Supplementary Table 61).

Analysis of time-related genes during cell type maturation

Early organ formation in mice begins at E10.5, and cells undergo the differentiation to reach maturity during development⁷⁷. Thus, we identified time-related genes that showed up-regulation patterns at the expression levels during the developmental processes by using the Spearman rank correlation analysis for different lineages in each tissue⁷⁸. Spearman rank correlation coefficient which has low requirements on data distribution and a high tolerance for outliers can directly reflect the monotonous relationship between variables, so we adopted

it. We treated the 7-stage information (E10.5Day, E12.5Day, E14.5Day, P0, P10, P21, and Adult) as the vectors labeled (1, 2, 3, 4, 5, 6, and 7), and then calculated the correlation between the gene expression levels across 7 development stages and the vectors for each stage. The larger the absolute value of the correlation coefficient is, the stronger the monotonicity of the gene expression level and time points is. The TFs with the Spearman rank correlation coefficient ≥ 0.8 in at least four lineages in five tissues with a p-value ≤ 0.05 were retained as the common time-upregulated TFs during lineage maturation.

Single-cell entropy analysis

Single-cell entropy estimation was performed using three methods: CCAT²⁹ (SCENT v1.0.2), SLICE²⁷ (v0.99.0), and StemID²⁶ (RaceID v0.2.2). To obtain the best performance, normalization was dependent on the computational methods. For CCAT, it is an approximation of network entropy. We applied CCAT to compute the correlations with the connectome and transcriptome based on the 'net13Jun12.m' PPI. We performed CCAT analysis by using a weighted matrix to leverage all the homology genes between human and other species. The weighted matrix was obtained by converting the gene homology relationship (one-to-one, one-to-many, many-to-one, and many-to-many) into a binary matrix and normalized it to one human gene. In StemID, it estimates the Shannon entropy of a cell's transcriptome directly based on the expression of each gene. We used StemID to infer entropies with default parameters. For SLICE, it established a kappa matrix of Gene Ontology annotations of the human or mouse to evaluate the probability distribution of the functional activation of each cell. SLICE was performed a deterministic calculation of scEntropy of individual cells over the Gene Ontology (GO) cluster activation profile with iter = 50. Cells were downsampled to 2,000 per tissue per stage to cut the calculation burden of SLICE and StemID. In summary, CCAT calculates the entropy-related values from the perspective of the network entropy of the gene interaction network. SLICE and StemID calculate the entropy values by using the activation of the gene pathway and the gene expression as probabilistic events, respectively. Although the principles of the three methods are different, their central idea is to coupled entropy with developmental potential. They evaluate biological systems using physical concepts and reflect the physical properties of biological systems.

Construction of a cell type hierarchy across species and gene regulation analysis

For invertebrates, to infer the topological relations of cell type development, we first constructed a PAGA graph⁴³ per lineage. We processed the data following the steps suggested by SCANPY, including total count normalization, log1p transformation, highly variable gene extraction, potential regression of confounding factors of genes and counts, scaling to z-scores, and PCA. Then, we computed a neighborhood graph among data points and used UMAP for topologically faithful embedding with min_dist = 0.1. Then, PAGA was performed with iter = 1,000. The cell type tree layout was based on a minimum spanning tree fitted to edges weighted by inverse connectivity. Edges in an abstracted graph with a probability higher than 0.0005 were considered as possible connections of cell-type hierarchies. For *S. mediterranea*, cell-type hierarchies were obtained from the consolidated lineage tree which was provided in paper¹² and for *C. intestinalis*, lineage and stage information were directly from the paper¹⁶. For complex vertebrates, we connected cell states across time according to gene expression similarity⁴⁴. For each tissue, we asked each adult cluster to 'vote' on its most likely ancestor

cluster from the fetal stage. To eliminate the influence of cell number, we randomly sampled 150 cells to embed them into the PCA space learned from the second time point only and kept nontrivial PCs as defined above. Then, in this embedding, for each cluster in the late time point, the cluster identities of the 5 nearest neighbors of each constituent cell from the previous time point were determined using a Euclidean distance metric. The percentages of votes cast for each possible ancestor were calculated, and the maximum frequencies of votes (20%-100%) of the cells in the cell group decided the ancestor group. For zebrafish datasets, we integrated the data using Seurat (v4.0.1) anchors integration functions to do the batch correction before the PCA analysis. Sankey plots were generated using the networkD3 (<https://christophergandrud.github.io/networkD3/>) R package. For atlas projects across species, we performed the same DGE analysis for cells in each tissue-cell type/lineage-cell type separately according to the cell type hierarchy using FindMarkers function in Seurat (v4.0.1). Wilcoxon rank sum test was performed to determine the statistical significance and benjamini-hochberg correction was used for the p-value adjustment. The top commonly differentially expressed genes (20%-100% of total cell types pairs, mean 60%-94% lineages, $p\text{-adj} < 0.1$, $\log_2\text{FC} \geq 0.25$, $\text{min_pct} \geq 0.1$) were estimated according to the frequency of differential expression in all unstable-to-stable cell-type pairs across species. To match the two time points (fetal and adult) of humans, only the E14.5 and adult stages of mice and 24-hpf and 3-month stages of zebrafish were considered for cross-species analysis. Genes that display consistent patterns in at least three species were defined as commonly upregulated and downregulated genes. Genes with either “up” or “down” regulated were excluded in the analysis.

The top 20 most negative TFs of the upregulated TFs were determined by Pearson correlations based on single-cell datasets and visualized by Cytoscape (v 3.5.0)⁷⁹.

Collection and prediction of orthologous genes and TFs

For *Homo sapiens*, *Mus musculus*, *Danio rerio*, and *C. elegans*, orthologous pairs were obtained from Ensembl v96 by BioMarkt. The transcriptome of *S. mediterranea* was downloaded from the PlanMine database⁸⁰ (*S. mediterranea*, dd_Smed_v6). The transcriptome of *H. vulgaris* was downloaded from the website <https://research.nhgri.nih.gov/hydra/download/?dl=tr>. The transcriptome of *C. intestinalis* was downloaded from http://ghost.zool.kyoto-u.ac.jp/download_kh.html. Then, the protein-coding sequence (CDS) was predicted by TransDecoder⁸¹ (v5.3.0) with the default parameters. Orthologous pairs were predicted by OrthoFinder⁸² (v2.2.6) with CDS files as the input. Here, we considered only one-to-one orthologous pairs with humans for commonly regulated genes. As for species-specific TFs, TFs of *Homo sapiens*, *Mus musculus*, *Danio rerio*, *C. elegans* were downloaded from AnimalTFDB 3.0 database⁸³. Other species-specific TFs except *H. vulgaris* were obtained from the paper⁵⁵. Genes from *H. vulgaris* were obtained with Swiss-Prot IDs of best hits. Thus, the TFs of *H. vulgaris* were defined by the genes annotated with the GO terms downloaded from the uniprot website: DNA-binding transcription factor activity or transcription factor binding. Those Swiss-Prot IDs of best hits were also checked for TFs from AnimalTFDB 3.0 and used as a supplement to TFs.

Lineage-specific TFs analysis across species

We applied two methods to calculate the lineage evolution relationship across species with the pseudo cell as inputs (aggregated every 20 cells from each cell type): SAMap⁸⁴ (v0.3.0) and MetaNeighbor⁸⁵ (pyMN v0.1.0). SAMap enables mapping single-cell transcriptomes between phylogenetically remote species based on the expression similarity while MetaNeighbor has high replicability in cell type matching using homologous weighted gene matrices. For SAMap, it constructs a gene-gene bipartite graph with cross-species edges connecting homologous gene pairs, weighted by protein sequence similarity. For MetaNeighbor, we constructed weighted matrices to leverage all the homology genes between humans and other species. The weighted matrices were obtained by converting the gene homology relationship (one-to-one, one-to-many, many-to-one, and many-to-many) into a binary matrix and normalized it to one human gene each. Lineage pairs with high confidence thresholds (alignment scores with > 0.5 in SAMap and Mean_AUROC > 0.8 in MetaNeighbor) were considered as highly reliable and biological plausible matches from different aspects. The combined projection of seven species was obtained from the function “SAMAP.scatter” from SAMap. The specificity of TFs was characterized using modified regulon specificity scores in SCENIC with TF expression count matrices as input^{54,55}. We then calculated the Z-score normalized TF specificity score to predict the essential TFs in each lineage. Development-related lineage-specific TFs were intersected with upregulated genes across species. The sequence similarity score was determined by NCBI BLAST with transcriptome or proteome data as inputs. An E-value threshold of $1e-6$ was set. It was also integrated into SAMap.

Pathway enrichment analysis

We used clusterProfiler⁸⁶ (v3.14.3) to perform Gene Ontology biological pathway enrichment analysis and orthologous genes were taken as the universe. Hypergeometric test was performed to identify significant go terms and benjamini-hochberg correction was used to adjust p values. We considered biological pathways with p adjusted values smaller than 0.05. We used REVIGO⁸⁷ to visualize the enrichment results. For Extended Data Fig. 1e, we used clusterProfiler to perform Gene Ontology biological pathway enrichment analysis for DEGs at neighboring stages. We considered biological pathways with $p\text{-adj} \leq 0.01$. For each stage, the enrichment terms as determined by clusterProfiler were used to manually combine into 13 ‘super terms’ for biological processes. For Extended Data Fig. 1i, GO enrichment analysis was first computed using the DEGs of the kidneys. Then, the enrichment scores of the terms were calculated and aggregated for each stage using AUCell³².

PPI analysis

We downloaded the PPI resource of human genes from STRING⁸⁸ (v11). Experimentally validated interactions from humans and transferred by homology from other species were used for the analysis. Then, we compared the log10 values of PPI number of four groups, the upregulated genes in at least three species, downregulated genes in at least three species, other conserved genes in at least three species, and all other genes in the PPI resource. We also downloaded the gene functional assignments from the eggNOG database (v5.0) and used the mammals’ non-supervised orthologous groups (maNOG) to assign genes into 26 categories⁸⁹. The 26 gene categories were arranged by their average number of PPI in ascending orders. Statistical analyses were done with R package ggpubr (v0.4.0,

<https://rpkg.sciencedata.com/ggpubr/>) for the two-tailed Wilcoxon rank sum test to determine the statistical significance of the differences between two groups.

Analysis of the CCAT driving gene across species

CCAT directly measures the correlation between transcriptome and connectome and therefore will be positive if the majority of network hubs are overexpressed in more potent cells²⁹. Thus, we used the number of adjacent edges to evaluate the degree of each gene in the PPI network and the top 20% of genes are regarded as network hubs. We intersected them with the commonly downregulated genes we found in the manuscript (highly expressed in undifferentiated cells, Supplementary Table 64) in each species as CCAT driving genes in more potent cells. Genes that appeared in at least five species were regarded as conserved CCAT driving genes. We performed gene enrichment analysis using clusterProfiler on those conserved CCAT driving genes. The biological processes related to ribosome biogenesis were marked red according to the previous paper⁵⁰.

Analysis of *Xbp1* expression pattern in MCDA

Given the low detection rate of transcriptional factors in the single-cell experiment, we chose the high-quality cells with more than 800 genes and calculated the average expression of *Xbp1* by normalized to a group of stably expressed gene sets generated from scMerge R package (v1.2.0, <https://bioconductor.org/packages/release/bioc/html/scMerge.html>). We used linear regression to measure the expression trend of *Xbp1* with a 95% confidence interval.

Cell type composition analysis

Significant differences in cell-type composition between groups were assessed using a propeller test from the speckle R package (v0.0.1, <https://github.com/Oshlack/speckle/>). We considered groups with FDR ≤ 0.01 to represent significantly changed cell types.

Gene expression variability analysis

To detect sensitive changes in weakly expressed genes, we calculated the gene expression variability using VarID⁷⁰ (RaceID, v0.2.2). We ran VarID with regNB=FALSE, k=10 for the pruning step, no_cores=10, and default parameters otherwise.

Analysis of global proteomics data

Liquid chromatography mass spectrometry (LC-MS) proteomic analysis was carried out by PTM Bio⁹⁰. Briefly, mouse embryos were ground into powder in liquid nitrogen and suspended in an ice-cold lysis buffer with 1% Triton X-100 and 1% protease inhibitor based on occasional sonication. The cell lysates were centrifuged at 12,000 g at 4°C for 15 min. The supernatants were collected and the protein concentration was measured. Proteins were precipitated using 20% trichloroacetic acid for 2 h at 4°C and then centrifuged at 4,500g for 5 min. The precipitate was washed three times with cold acetone. The dried protein pellets were resuspended within 200 mM tetraethylammonium bromide (TEAB) based on occasional sonication and then digested with trypsin overnight. Dithiothreitol (DTT) was added to a final concentration of 5 mM and the supernatants were incubated at 56°C for 30 min. Iodoacetamide (IAA) was added to a final concentration of 11 mM and the supernatants were incubated in the dark for 15 min. Peptides were separated using NanoElute and analyzed using timsTOF Pro. The resulting

MS/MS data were processed using the MaxQuant search engine (v.1.5.2.8, <https://www.maxquant.org/>) and mapped to the Mus_musculus_10090 database. The FDR was adjusted to < 1%, and the minimum score for modified peptides was set to > 40. Trypsin/P was defined as the cleavage enzyme, and up to two missing cleavages were allowed. For proteomic analysis, the first search range was set to 5 ppm for precursor ions, and the main search range was set to 5 ppm and 0.02 Da for fragment ions. Carbamidomethylation of cysteines was defined as the fixed modification, and oxidation on methionine was defined as the variable modification. The quantification method used was LFQ, the FDR was adjusted to < 1%, and the minimum score for modified peptides was > 40.

scATAC-seq and ChIP-seq data analysis

We used the ChromVAR⁹¹ (v1.12.0) to calculate the accessibility of the *Xbp1* motif in scATAC-seq datasets to compare the *Xbp1* motif enrichment between differentiated states and undifferentiated states both in the human and mouse. The mouse scATAC-seq data was downloaded from two papers^{61,61} and human scATAC-seq data from another paper⁶³. The motif PWM was downloaded from the CisBP database (<http://cisbp.cabr.utoronto.ca/>). For better visualization, we arranged the cells according to their differentiated states. This comparison was restricted with the cell-type annotations provided. As shown, the *Xbp1* motif was less opened in undifferentiated cells in both human and mouse tissues in neuron cell types and hemopoietic cell types. ChIP-seq data for *Xbp1* were downloaded from previous studies^{58,69,92,93}. The target genes were binarized and integrated for visualization.

Data Availability

The data generated in this study can be downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under accession numbers GSE176063 and GSE178217. We also provide an interactive website (<http://bis.zju.edu.cn/MCA/>) to enable public access to the MCDA data.

The following publicly available datasets were used in the study: Mus_musculus. GRCm38.88 genome, Mus_musculus_10090 database, AnimalTFDB 3.0 database, STRING database (v11), eggNOG database (v5.0), Ensembl v96; the *Schmidtea mediterranea* dataset generated by Plass et al. (GSE103633), the *Caenorhabditis elegans* dataset generated by Packer et al. (GSE126954); the *Ciona intestinalis* dataset generated by Cao et al. (GSE131155); the *Hydra vulgaris* dataset generated by Siebert et al. (GSE121617); the *Danio rerio* dataset generated by Li et al. (GSE178151); the *Homo sapiens* dataset generated by Han et al. (GSE134355), and part of *Mus musculus* dataset (E14.5 and adult) generated by Han et al. (GSE108097 and GSE134355). The mouse scATAC-seq dataset generated by Cusanovich et al (<https://atlas.gs.washington.edu/mouse-atac/data/>) and Di Bella et al. (GSE153164), and the human scATAC-seq dataset generated by Domcke et al. (descartes.brotmanbaty.org).

Code Availability

Detailed code is available at GitHub (<https://github.com/ggjlabs/MCDA/>) and Zenodo (<https://zenodo.org/record/6423564#.Yk-YgShBw2w>) (ref.⁹⁴).

989

990 **Statistics & Reproducibility**

991 No statistical methods used to predetermine sample size. 520,801 single cells were analyzed in
992 total for a time-series mouse cell differentiation atlas construction. A total of 52 mouse tissues
993 from different development stages were analyzed. 2-4 replications were done for different
994 tissues. The results of major cell type clusters are reproducible. Experimental mice and
995 embryos were randomized before sample preparation. Different single cells were randomly
996 captured before analysis. For all experiments, investigators were blinded to group allocation
997 during the data collection and analysis. All related statistical methods and sample size are
998 described in the figure legends and the Method section.

999

1000 **Method reference**

1001 71. Xie, Y. *et al.* An episomal vector-based CRISPR/Cas9 system for highly efficient gene knockout in human
1002 pluripotent stem cells. *Scientific reports* **7**, 1–11 (2017).

1003 72. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

1004 73. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis.
1005 *Genome Biology* **19**, (2018).

1006 74. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene
1007 expression data. *Nature biotechnology* **33**, 495–502 (2015).

1008 75. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science*
1009 **367**, eaay3224 (2020).

1010 76. Zhou, Q. *et al.* A mouse tissue transcription factor atlas. *Nature communications* **8**, 1–15 (2017).

1011 77. Lambert, L. J., Muzumdar, M. D., Rideout III, W. M. & Jacks, T. Basic Mouse Methods for Clinician
1012 Researchers: Harnessing the Mouse for Biomedical Research. in *Basic Science Methods for Clinical*
1013 *Researchers* 291–312 (Elsevier, 2017).

1014 78. Teschendorff, A. E. & Wang, N. Improved detection of tumor suppressor events in single-cell RNA-Seq
1015 data. *npj Genom. Med.* **5**, 43 (2020).

1016 79. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nature methods* **9**, 1069–1076 (2012).

1017 80. Rozanski, A. *et al.* PlanMine 3.0—improvements to a mineable resource of flatworm biology and
1018 biodiversity. *Nucleic acids research* **47**, D812–D820 (2019).

1019 81. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for
1020 reference generation and analysis. *Nature protocols* **8**, 1494–1512 (2013).

1021 82. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons
1022 dramatically improves orthogroup inference accuracy. *Genome biology* **16**, 1–14 (2015).

1023 83. Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal
1024 transcription factors. *Nucleic acids research* **47**, D33–D38 (2019).

1025 84. Tarashansky, A. J. *et al.* Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife*
1026 **10**, e66747 (2021).

1027 85. Fischer, S., Crow, M., Harris, B. D. & Gillis, J. Scaling up reproducible research for single-cell
1028 transcriptomics using MetaNeighbor. *Nature Protocols* 1–37 (2021).

1029 86. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes
1030 among gene clusters. *Omics: a journal of integrative biology* **16**, 284–287 (2012).

1031 87. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene
1032 ontology terms. *PloS one* **6**, e21800 (2011).

1033 88. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage,
1034 supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**, D607–
1035 D613 (2019).

1036 89. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and
1037 annotation. *Nucleic acids research* **49**, W293–W296 (2021).

1038 90. Song, Y. *et al.* Screening of Potential Biomarkers for Gastric Cancer with Diagnostic Value Using Label-
1039 free Global Proteome Analysis. *Genomics, Proteomics & Bioinformatics* (2021).

- 1040 91. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-
1041 associated accessibility from single-cell epigenomic data. *Nature Methods* **14**, 975–978 (2017).
- 1042 92. Argemí, J. *et al.* X-box binding protein 1 regulates unfolded protein, acute-phase, and DNA damage
1043 responses during regeneration of mouse liver. *Gastroenterology* **152**, 1203-1216. e15 (2017).
- 1044 93. Khetchoumian, K. *et al.* Pituitary cell translation and secretory capacities are enhanced cell autonomously
1045 by the transcription factor Creb3l2. *Nature communications* **10**, 1–13 (2019).
- 1046 94. Fei, L. ggjlab/MCDA: v1.0.0. (2022) doi:10.5281/zenodo.6423564.

1047

1048

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [InventoryofSupplementaryInformation.pdf](#)