Mouse lemur transcriptomic atlas elucidates primate genes, physiology, disease, and evolution

The Tabula Microcebus Consortium*

The Tabula Microcebus Consortium

Lead authors

Camille Ezran[1, 2] ‡, Shixuan Liu[1, 2, 3] ‡, Jingsi Ming[4, 5], Lisbeth A. Guethlein[6, 7], Michael F.Z. Wang[8, 9], Roozbeh Dehghannasiri[1, 10], Julia Olivieri[11], Hannah K. Frank[12, 13], Alexander Tarashansky[14, 15], Winston Koh[16, 17], Qiuyu Jing[18], Olga Botvinnik[15], Jane Antony[19], Stephen Chang[1, 2, 20], Angela Oliveira Pisco[15], Jim Karkanias[15], Can Yang[5], James E. Ferrell, Jr.[1, 3], Scott D. Boyd[12], Peter Parham[6, 7], Jonathan Z. Long[12, 21], Bo Wang[14], Julia Salzman[1, 10], Iwijn De Vlaminck[8], Angela Wu[18, 22, 23], Stephen R. Quake[14, 15, 24]†, Mark A. Krasnow[1, 2] †

‡ Co-first authors

† Corresponding authors: krasnow@stanford.edu (M.A.K.), steve@quake-lab.org (S.R.Q.)

Leadership and coordination

Trainees: Camille Ezran[1, 2], Shixuan Liu[1, 2, 3]

Principal investigators: Mark A. Krasnow[1, 2], Stephen R. Quake[14, 15, 24], Angela Wu[18, 22, 23], Iwijn De Vlaminck[8], Bo Wang[14], Julia Salzman[1, 10], Jim Karkanias[15], Angela Oliveira Pisco[15]

Donor recruitment and animal husbandry

Megan A. Albertelli[25], Caitlin J. Karanewsky[1, 2], Jozeph L. Pendleton[1, 2], Fabienne Aujard[26], Martine Perret[26], Liza Shapiro[27], Andriamahery Razafindrakoto[28], Hajanirina Noëline Ravelonjanahary[28], Patricia Wright[29], Anne D. Yoder[30], Cathy V. Williams[31], Robert Schopler[31], Ute Radespiel[32], Jean-Michel Verdier[33], Corinne Lautier[33], E. Christopher Kirk[27], Rebecca Lewis[27]

Necropsy, sample collection, biobanking, histology, and pathology

Kerriann M. Casey[25], Megan A. Albertelli[25], Caitlin J. Karanewsky[1, 2], Jozeph L. Pendleton[1, 2], Camille Ezran[1, 2], Shixuan Liu[1, 2, 3], Kyle J. Travaglini[1, 2], Astrid Gillich[1, 2], Zicheng Zhao[2, 25], Stephen Chang[1, 2, 20], Elias Godoy[25], Jérémy Terrien[26], Jacques Epelbaum[26, 34], Dita Gratzinger[12], Katherine Lucot[12], Tom Montine[12]

Tissue processing, annotation, and analysis [Principal investigators]

Aorta: Jessica D'Addabbo[35, 36], Isaac Bakerman[20], [Patricia Nguyen[35, 36, 37]]

Bladder: Aaron Kershner[1, 38], Karim Mrouj[19], [Philip Beachy[3, 19, 39, 40]]

Blood: Rahul Sinha[41], Camille Ezran[1, 2], Yue Zhang[2, 42], Shixuan Liu[1, 2, 3], Caitlin J. Karanewsky[1, 2], [Irving L. Weissman[41], Mark A. Krasnow[1, 2]]

Bone (limb, spine): Thomas Ambrosi[19], Malachia Hoover[19], Alina Alam[19], [Charles Chan[19]]

Bone marrow: Rahul Sinha[41], SoRi Jang[1, 2], Camille Ezran[1, 2], [Irving L. Weissman[41], Mark A. Krasnow[1, 2]]

Brain (cortex, brainstem, cerebellum): Avin Veerakumar[1, 2, 14, 43], Peng Li[1, 2], Andrea R. Yung[1, 2], Connor V. Duffy[2, 44], Song-Lin Ding[45], [Ed S. Lein[45], Silvana Konermann[1, 2], Liqun Luo[2, 42], Trygve E. Bakken[45], Justus Kebschull[46], Rebecca D. Hodge[45], Mark A. Krasnow[1, 2]]

Colon/Small intestine: Taichi Isobe[47], [Michael F. Clarke[19]]

Diaphragm/Limb muscle: Antoine de Morree[48, 49], Biter Bilen[48], Jean Farup[48, 49], Andoni Urtasun[48], Jengmin Kang[48], [Thomas A. Rando[48]]

Eye: Ming Chen[44], Bao Xiang[50], Varun Ramanan Subramaniam[50], Shravani Mukherjee[50], Aditi Swarup[50], Lily Kim[44], Bronwyn Scott[50], Ahmad Al-Moujahed[50], [Albert Y. Wu[50], Douglas Vollrath[44], Lubert Stryer[51]]

Fat: Nicholas Schaum[48], Jingsi Ming[4, 5], Shixuan Liu[1, 2, 3], Amanda L. Wiggenhorn[12, 52], [Tony Wyss-Coray[48, 53], Jonathan Z. Long[12, 21], Angela Wu[18, 22, 23]]
Ganglion (vagus nerve): Yin Liu[1, 2], [Mark A. Krasnow[1, 2]]
Heart: Stephen Chang[1, 2, 20], [Mark A. Krasnow[1, 2]]
Hypothalamus & Pituitary: Andrea R. Yung[1, 2], Shixuan Liu[1, 2, 3], [Mark A. Krasnow[1, 2]]
Kidney: Shixuan Liu[1, 2, 3], Ahmad Nabhan[1, 2], Lolita Penland[15], Yue Zhang[2, 42], Andrea R. Yung[1, 2], [Gabriel Loeb[54], Mark A. Krasnow[1, 2]]
Liver: Shengda Lin[55], Shixuan Liu[1, 2, 3], Honor Paine[56], Deviana Burhan[56], Aris Taychameekiatchai[56], [Steven Artandi[1, 37, 57], Bruce Wang[56]]
Lung: Kyle J. Travaglini[1, 2], Ahmad Nabhan[1, 2], F. Hernán Espinoza[1, 2], Astrid Gillich[1, 2], [Christin Kuo[58], Ross Metzger[58, 59], Mark A. Krasnow[1, 2]]
Lymph node: SoRi Jang[1, 2], Lolita Penland[15], [Norma Neff[15], Mark A. Krasnow[1, 2]]
Mammary gland: Zhen Qi[19], [Michael F. Clarke[19]]
Microbiome: Rebecca Culver[44], [Kerwyn C. Huang[7, 14, 15]]
Pancreas: Patrick Neuhöfer[1, 37, 57], Charles A. Chang[39, 60], Yan Hang[39, 60], [Seung K. Kim[37, 39, 60, 61], Steven Artandi[1, 37, 57]]
Prostate: Hannah Weinstein[62, 63, 64], Paul Allegakoen[62], [Franklin W. Huang[62]]
Salivary gland: Sivakamasundari V[19], [Philip Beachy[3, 19, 39, 40]]
Skin: Song Eun Lee[48, 53], [Tony Wyss-Coray[48, 53]]
Spleen: Lolita Penland[15], Hannah K. Frank[12, 13], Dita Gratzinger[12], [Norma Neff[15], Scott D. Boyd[12]]
Testes: Kazuteru Hasegawa[1, 37, 57], Hosu Sin[39], Jingsi Ming[4, 5], [Steven Artandi[1, 37, 57], Margaret Fuller[39, 44], Angela Wu[18, 22, 23]]
Thymus: Isaac Bakerman[20], Jessica D'Addabbo[35, 36], [Patricia Nguyen[35, 36, 37]]
Tongue: Wan-Jin Lu[19], Ankit Baghe[l42], [Philip Beachy[3, 19, 39, 40]]
Trachea/Thyroid/Parathyroid: William Kong[19], [Philip Beachy[3, 19, 39, 40]]
Uterus/Abdominal mass: Lolita Penland[15], [Norma Neff[15]]

Library processing, sorting, and sequencing
Stephen Chang[1, 2, 20], Lolita Penland[15], Carly Israel[15], Rene Sit[15], Jennifer Okamoto[15], Yue Zhang[2, 42], Rahul Sinha[41], Shixuan Liu[1, 2, 3], Camille Ezran[1, 2], Kyle J. Travaglini[1, 2], Antoine de Morree[48, 49], Ahmad Nabhan[1, 2], Youcef Ouadah[1, 65], Jalal Baruni[1, 2, 42, 66], Robert C. Jones[14], Norma Neff[15], Maurizio Morri[15], Spyros Darmanis[15], Sheela Crasta[15], Jia Yan[15], Aditi Agrawal[15], Shelly Huynh[15], Brian Yu[15], James Webber[15]

Global data analysis [Principal investigators]
General analysis: Camille Ezran[1, 2], Shixuan Liu[1, 2, 3], [Mark A. Krasnow[1, 2]]
Sequence alignment: Stephen Chang[1, 2, 20], Kyle J. Travaglini[1, 2], Rahul Sinha[41], Rene Sit[15], [Mark A. Krasnow[1, 2], Irving L. Weissman[41], Stephen R. Quake[14, 15, 24]]
FIRM integration: Jingsi Ming[4, 5], [Angela Wu[18, 22, 23], Can Yang[5]]
Molecular trajectory: Shixuan Liu[1, 2, 3], Camille Ezran[1, 2], Weilun Tan[7], [Mark A. Krasnow[1, 2]]
Global cell type comparison: Shixuan Liu[1, 2, 3], [Mark A. Krasnow[1, 2]]
Cross-species comparison: Olga Botvinnik[15], Alexander Tarashansky[14, 15], Saba Nafees[15], Shixuan Liu[1, 2, 3], Camille Ezran[1, 2], Kyle J. Travaglini[1, 2], Antoine de Morree[48, 49], Jia Zhao[5], Gefei Wang[5], [Jim Karkanias[15], Bo Wang[14], Stephen R. Quake[14, 15, 24], Mark A. Krasnow[1, 2], Thomas A. Rando[48], Angela Wu[18, 22, 23], Can Yang[5]]

Spermatogenesis: Jingsi Ming[4, 5], Hosu Sin[39], Kazuteru Hasegawa[1, 37, 57], Camille Ezran[1, 2], [Angela Wu[18, 22, 23], Margaret Fuller[39, 44], Steven Artandi[1, 37, 57], Can Yang[5], Mark A. Krasnow[1, 2]]
uTAR: Michael F.Z. Wang[8, 9], [Iwijn De Vlaminck[8]]
SICILIAN: Roozbeh Dehghannasiri[1, 10], Julia Olivieri[11], [Julia Salzman[1, 10]]
SAMap: Alexander Tarashansky[14, 15], [Bo Wang][14]
Major histocompatibility locus: Lisbeth A. Guethlein[6, 7], Camille Ezran[1, 2], [Peter Parham[6, 7], Mark A. Krasnow[1, 2]]
B cell receptor locus: Hannah K. Frank[12, 13], Camille Ezran[1, 2], [Scott D. Boyd[12], Mark A. Krasnow[1, 2]]
Cross-tissue neutrophil trajectory: Camille Ezran[1, 2], Shixuan Liu[1, 2, 3], [Mark A. Krasnow[1, 2]]
Cross-tissue monocyte and macrophage comparison: Qiuyu Jing[18], Camille Ezran[1, 2], Shixuan Liu[1, 2, 3], [Angela Wu[18, 22, 23], Mark A. Krasnow[1, 2]]
Chemokine expression: Camille Ezran[1, 2], [Mark A. Krasnow[1, 2]]
Metastatic tumor: Jane Antony[19], Camille Ezran[1, 2], [Michael F. Clarke[19], Mark A. Krasnow[1, 2]]
Endothelium: Camille Ezran[1, 2], Astrid Gillich[1, 2], Geoff Stanley[14], [Ross Metzger[58, 59], Mark A. Krasnow[1, 2], Stephen R. Quake[14, 15, 24]]
Adipocytes: Jingsi Ming[4, 5], Jinxurong Yang[18], Camille Ezran[1, 2], Shixuan Liu[1, 2, 3], [Jonathan Z. Long[12, 21], Angela Wu[18, 22, 23], Mark A. Krasnow[1, 2], Can Yang[5]]
Primate-specific genes: Winston Koh[16, 17], Camille Ezran[1, 2], [Mark A. Krasnow[1, 2]]
Global hormone signaling: Shixuan Liu[1, 2, 3], [James E. Ferrell, Jr.[1, 3], Mark A. Krasnow[1, 2]]
OnClass: [Sheng Wang[67]]
Data portal/Online histology atlas: Zicheng Zhao[2, 25], Snigdha Agarwal[15], Kyle Awayan[15], Aaron McGeever[15], Venkata N. P. Vemuri[15], [Stephen R. Quake[14, 15, 24], Angela Oliveira Pisco[15]]

Principal investigators (A-Z)
Megan A. Albertelli[25], Steven Artandi[1, 37, 57], Fabienne Aujard[26], Trygve E. Bakken[45], Philip Beachy[3, 19, 39, 40], Scott D. Boyd[12], Kerriann M. Casey[25], Charles Chan[19], Michael F. Clarke[19], Jacques Epelbaum[26, 34], James E. Ferrell, Jr.[1, 3], Margaret Fuller[39, 44], Dita Gratzinger[12], Franklin W. Huang[62], Kerwyn C. Huang[7, 14, 15], Jim Karkanias[15], Justus Kebschull[46], Seung K. Kim[37, 39, 60, 61], Silvana Konermann[1, 2], Mark A. Krasnow[1, 2], Christin Kuo[58], Ed S. Lein[45], Shengda Lin[55], Gabriel Loeb[54], Jonathan Z. Long[12, 21], Liqun Luo[2, 42], Ross Metzger[58, 59], Jingsi Ming[4, 5], Tom Montine[12], Norma Neff15, Patricia Nguyen[35, 36, 37], Peter Parham[6, 7], Martine Perret[26], Angela Oliveira Pisco[15], Stephen R. Quake[14, 15, 24], Thomas A. Rando[48], Ute Radespiel[32], Julia Salzman[1, 10], Liza Shapiro[27], Lubert Stryer[51], Jérémy Terrien[26], Jean-Michel Verdier[33], Iwijn De Vlaminck[8], Douglas Vollrath[44], Bo Wang[14], Bruce Wang[56], Sheng Wang[67], Irving L. Weissman[41], Patricia Wright[29], Tony Wyss-Coray[48, 53], Albert Y. Wu[50], Angela Wu[18, 22, 23], Can Yang[5], Anne D. Yoder[30]

Writing
Camille Ezran[1, 2], Shixuan Liu[1, 2, 3], Jingsi Ming[4, 5], Lisbeth A. Guethlein[6, 7], Michael F.Z. Wang[8, 9], Roozbeh Dehghannasiri[1, 10], Julia Olivieri[11], Hannah K. Frank[12, 13], Alexander Tarashansky[14, 15], Qiuyu Jing[18], Jane Antony[19], Peter Parham[6, 7], Jonathan Z. Long[12, 21], Bo Wang[14], Iwijn De Vlaminck[8], Angela Wu[18, 22, 23], Stephen R. Quake[14, 15, 24], Mark A. Krasnow[1, 2]

Affiliations
1. Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA.
2. Howard Hughes Medical Institute, USA.
3. Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA.
4. Academy of Statistics and Interdisciplinary Sciences, KLATASDS-MOE, East China Normal University, Shanghai, China.
5. Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong SAR, China.
6. Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, USA.
7. Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.
8. Nancy E. and Peter C. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA.
9. Department of Computational Biology, Cornell University, Ithaca, NY, USA.
10. Department of Biomedical Data Science, Stanford University, Stanford, CA, USA.
11. Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA.
12. Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA.
13. Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA.
14. Department of Bioengineering, Stanford University, Stanford, CA, USA.
15. Chan Zuckerberg Biohub, San Francisco, CA, USA.
16. Institute of Bioengineering and Bioimaging, Agency of Science Technology and Research, Singapore
17. Bioinformatics Institute, Agency of Science Technology and Research, Singapore
18. Division of Life Science, Hong Kong University of Science and Technology, Hong Kong SAR, China.
19. Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA.
20. Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA.
21. Stanford ChEM-H, Stanford, CA, USA.
22. Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Hong Kong SAR, China.
23. Center for Aging Science, Hong Kong University of Science and Technology, Hong Kong SAR, China.
24. Department of Applied Physics, Stanford University, Stanford, CA, USA.
25. Department of Comparative Medicine, Stanford University School of Medicine, Stanford, CA, USA.
26. Adaptive Mechanisms and Evolution (MECADEV), UMR 7179, National Center for Scientific Research, National Museum of Natural History, Brunoy, France.
27. Department of Anthropology, University of Texas at Austin, Austin, TX, USA.

28. Department of Animal Biology, Faculty of Science, University of Antananarivo, Antananarivo 101, BP 566, Madagascar.
29. Department of Anthropology, Stony Brook University, NY, USA.
30. Department of Biology, University Program in Genetics & Genomics, Duke University, Durham, NC, USA.
31. Duke Lemur Center, Durham, NC, USA.
32. Institute of Zoology, University of Veterinary Medicine Hannover, Germany.
33. MMDN, University of Montpellier, EPHE-PSL, INSERM, Montpellier, France.
34. Unité Mixte de Recherche en Santé 894 INSERM, Centre de Psychiatrie et Neurosciences, Université Paris Descartes, Sorbonne, Paris, France
35. Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA.
36. Stanford Cardiovascular Institute, Stanford, CA, USA.
37. Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.
38. Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA.
39. Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA.
40. Department of Urology, Stanford University School of Medicine, Stanford, CA, USA.
41. Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford, CA, USA.
42. Department of Biology, Stanford University, Stanford, CA, USA.
43. Medical Scientist Training Program, Stanford University School of Medicine, Stanford, CA, USA.
44. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.
45. Human Cell Types Department, Allen Institute for Brain Science, Seattle, WA, USA.
46. Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD, USA.
47. Department of Oncology and Social Medicine, Kyushu University, Japan.
48. Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA.
49. Department of Biomedicine, Aarhus University, Aarhus, Denmark.
50. Department of Ophthalmology, Stanford University School of Medicine, Stanford, CA, USA.
51. Department of Neurobiology, Stanford University School of Medicine, Stanford, CA, USA.
52. Department of Chemistry, Stanford University, Stanford, CA, USA.
53. Wu Tsai Neurosciences Institute, Stanford, CA, USA.
54. Division of Nephrology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA.
55. Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang, China.
56. Department of Medicine and Liver Center, University of California San Francisco, San Francisco, CA, USA.
57. Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA.
58. Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA.
59. Division of Cardiology, Stanford University School of Medicine, Stanford, CA, USA.
60. Stanford Diabetes Research Center, Stanford, CA, USA.
61. JDRF Center of Excellence, Stanford, CA, USA.

62. Division of Hematology/Oncology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA.
63. Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA.
64. Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, USA.
65. Program in Cancer Biology, Stanford University School of Medicine, Stanford, CA, USA.
66. Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA, USA.
67. Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, USA.

ABSTRACT

Mouse lemurs (*Microcebus* spp.) are an emerging model organism for primate biology, behavior, health, and conservation. Although little has been known about their cellular and molecular biology, in the accompanying paper we used large-scale single cell RNA-sequencing of 27 organs and tissues to identify over 750 molecular cell types and their full transcriptomic profiles. Here we use this extensive transcriptomic dataset to uncover thousands of previously unidentified genes and hundreds of thousands of new splice junctions in the reference genome that globally define lemur gene structures and cell-type selective expression and splicing and to investigate gene expression evolution. We use the atlas to explore the biology and function of the lemur immune system, including the expression profiles across the organism of all MHC genes and chemokines in health and disease, and the mapping of neutrophil and macrophage development, trafficking, and activation, their local and global responses to infection, and primate-specific aspects of the program. We characterize other examples of primate-specific physiology and disease such as unique features of lemur adipocytes that may underlie their dramatic seasonal rhythms, and spontaneous metastatic endometrial cancer that models the human gynecological malignancy. We identify and describe the organism-wide expression profiles of over 400 primate genes missing in mice, some implicated in human disease. Thus, an organism-wide molecular cell atlas and molecular cell autopsies can enhance gene discovery, structure definition, and annotation in a new model organism, and can identify and elucidate primate-specific genes, physiology, diseases, and evolution.

INTRODUCTION

Although many of the genes, pathways, and principles of modern biology and the molecular foundations of medicine were uncovered in studies of the canonical genetic model organisms such as yeast, *Drosophila*, and mouse, new model organisms are being developed to study aspects of biology and medicine not observed or poorly modeled in the canonical model organisms[1–5]. The recent explosion of these emerging model organisms in diverse areas of biology has been fueled by advances in DNA sequencing and genomics, which has made construction of a high quality reference genome for a newly studied species readily attainable, and by CRISPR-Cas9 systems that have made introduction of targeted mutations in the genome practical in many species, enabling genetic investigations. What has remained challenging is obtaining a rich cellular and molecular understanding of a new model organism and its biology, which has generally required decades to accrue through intensive effort of many researchers developing precision technologies and applying them to various aspects of the organism's development, physiology, and behavior. We reasoned that single cell transcriptomics performed at organism-wide scale could greatly facilitate such an understanding, and in the accompanying manuscript (Tabula Microcebus)[6] we created a comprehensive molecular cell atlas of the gray mouse lemur, *Microcebus murinus*.

Mouse lemurs (*Microcebus spp.*) provide an appealing model for the study of primate biology and health[7–9]. Their physiology has been studied for decades in laboratory colonies, especially their circadian and dramatic seasonal rhythms, metabolism, cognition and ageing[10–16]. Likewise, their ecology, behavior, communication, and phylogeny have been investigated through field studies in their native Madagascar[17–21]. However, little has been known about their genetics, genomics, or cellular and molecular biology. That began to change with the completion

of an annotated reference genome sequence for *M. murinus*[22]. And, in the accompanying paper (Tabula Microcebus)[6], we used large-scale droplet (10x) and plate-based (Smart-seq2, SS2) single cell RNA-sequencing (scRNA-seq) of 226,000 cells from 27 mouse lemur organs and tissues opportunistically procured from four aged donors that were clinically and histologically characterized. This identified over 750 molecular cell types and their full transcriptomic profiles, and defined the molecular relationships of the cell types to each other and to the homologous cell types in human and mouse.

In this paper, we use this molecular cell atlas and extensive transcriptomic dataset to elucidate primate genes, physiology, and disease. We uncover thousands of previously unidentified genes and hundreds of thousands of new splice junctions that globally define lemur gene structures and cell-type selective expression and splicing, and we show how these global expression profiles aid in gene annotation and characterizing gene evolution. We use the atlas to probe the genes, cells, and function of the lemur immune system in health and disease, including primate-specific genes and aspects of immune function. We characterize other examples of primate-specific physiology and disease, along with the expression profiles of over 400 primate genes absent in mice. The results show how a molecular cell atlas can be used to elucidate primate genes, physiology, diseases, and evolution, and the approach can be readily applied to other emerging model organisms.

RESULTS

1. Organism-wide scRNA-seq uncovers new genes

Our organism-wide scRNA-seq dataset provides a massive amount of transcriptomic sequence information: ~$2x10^{12}$ base pairs (~10^{12} bp high-quality reads from 10x and SS2 each) distributed throughout the ~$2.5x10^{9}$ bp *M. murinus* genome, providing ~10^{4}-fold average

coverage of the transcriptome (~$2.5x10^8$ bp of NCBI-annotated transcripts). We reasoned that such deep coverage of the transcriptome across most cell types of most organs would enhance gene detection, gene structure (intron-exon) definition[23], and gene annotation beyond current state-of-the-art methods that rely primarily on phylogenetic sequence comparisons and, when available, bulk RNA sequencing, such as the current *M. murinus* genome annotations (NCBI Microcebus Murinus Annotation Release 101, Ensembl Genome Browser v100) that draw on bulk RNA-seq data from eight tissues.

To evaluate the impact of our scRNA-seq datasets on gene detection, we used the hidden Markov model approach described in an associated paper[24] to identify Transcriptionally Active Regions (TARs) of the genome, genomic locations with significant read coverage, using the 10x generated sequencing reads from our scRNA-seq datasets (Fig. 1a). This revealed that TARs comprise 13% ($3.3x10^8$ bp) of the Mmur 3.0 genome assembly ($2.5x10^9$ bp), with most TARs (87%, $2.8x10^8$ bp, 11% of genome) mapping to previously annotated genes (aTARs) (Fig. 1b). The rest of the TARs (13%, $4.2x10^7$ bps, 1.7% of genome) mapped to unannotated genomic regions (uTARs), including newly identified genes (see below). uTARs that are differentially-expressed across cell types (see Methods) account for 2.4±1.5% (mean±s.d.) of the unique sequencing reads per cell, up to a median value of 18.5% in skin sweat gland cells (Fig. 1c, Table 1). These differentially-expressed uTARs are likely biologically significant because, in many tissues, from uTAR cellular expression patterns alone, we could distinguish cell types with a consistency that approached that using previously annotated genes or aTARs, as defined by the computed silhouette values (Fig. S1a-c). Likewise, dimensional reduction of the testes scRNA-seq dataset using uTAR features alone was able to reconstruct the full spermatogenesis program, mimicking the result with annotated genes[6] including stage-appropriate expression of major

sperm cell type marker genes and the decline during differentiation in the number of genes expressed (Fig. 1d-g, Fig. S1d).

We sought to determine the gene identities and molecular functions of uTARs. We first showed that TAR analysis has high sensitivity for detecting previously annotated genes (Fig. S1e). For example, of the 5000 NCBI-annotated mouse lemur genes with the highest cell type expression variance in our 10x scRNA-seq dataset, TARs captured 98% (4884 genes), and of the 3904 genes annotated by Ensembl but not NCBI, TARs captured 44% (1728 genes, mostly non-protein coding genes, Table 1). TARs also captured 88% (379 genes) of the 429 primate-specific genes (Section 7) (Table 7). We then searched for homologs in other species of the 4003 differentially-expressed uTARs (Table 1). DNA sequence (blastn) searches identified homologous genes for 2368 (59%) of these 4003 uTARs and transcript/protein sequence searches (nf-predictorthologs[25,26]) identified protein-coding (DIAMOND blastp) hits for 3185 (80%), noncoding (Infernal cmscan) hits for 45 (1%), and both protein-coding and noncoding hits for 231 (6%) (Fig. S1f). Some of these uTARs are conserved genes missed by the conventional annotation pipeline, such as dozens of genes present in human (with or without a mouse ortholog) but missing from NCBI and Ensembl's annotations of Mmur 3.0 (e.g. *DRD5* in fibroblasts of the bone, *GSTA3* in tendon cells of the bone, *EEFSEC* in pericytes of the hypothalamus, *MYEOV* in astrocytes of the brain stem, *TIGD1* in *CD4*-positive cells, *SPRR2G* in suprabasal epidermal cells, see Table 1).

We also used our scRNA-seq datasets to aid discovery of immunoglobulin (Ig) genes (Fig. 1h), which are exceedingly difficult to annotate by conventional approaches because they comprise large arrays of related and rapidly evolving genes and gene segments, some extremely short (e.g. D-region segments are as short as 10 bp) but widely spaced in the genome, which are

brought together by V(D)J recombination during B lymphocyte development to create antibody diversity[27]. We first used BLAST to search the mouse lemur genome (Mmur 3.0) for homologs of human Ig constant region sequences that are missing in the current annotation (NCBI Microcebus Murinus Annotation Release 101), which identified the *IGG* constant region. We then mapped to the genome the expressed Ig sequences from the 829 B lymphocytes and plasma cells in our SS2 dataset, which was especially valuable for this application because SS2 profiles contain sequence reads distributed across the full transcription unit. The mapped expressed sequences revealed the correct *IGA* and *IGM* gene structures, uncovered the D and J gene clusters, tripled the number of identified V genes (from 32 to 92), and suggested 15 V genes may be pseudogenes because they were not among the expressed genes. Some of the expressed V genes mapped to a large V gene cluster located ~5 MB upstream of the rest of the *IGH* locus, suggesting it is an orphan V gene cluster[28,29] or is incorrectly placed in the current genome assembly. This atlas-enhanced genome annotation shows that mouse lemur Ig heavy chain locus has a similar overall organization to that of human, but it is substantially streamlined with only a single constant region of each heavy chain isotype and no *IGD*, an evolutionarily plastic isotype lost in many lineages[30,31]. Hence mouse lemur may provide a simplified primate model for understanding immunoglobulin gene rearrangement, expression, and function, and B cell responses. Similar analyses of Ig light chain genes corrected the structure of *IGK* (kappa chain) and *IGL* (lambda) constant regions, uncovered duplicated *IGK* J gene clusters and additional *IGL* JC clusters, and doubled the number of identified V genes at each locus (from 26 to 59 genes at *IGK*, 28 to 61 at *IGL*), some of which may be pseudogenes. Thus, organism-wide scRNA-seq is a powerful way of detecting missed genes throughout the genome including historically challenging genomic regions like Ig loci.

2. Organism-wide scRNA-seq defines gene structures and splicing

To evaluate the impact of our scRNA-seq datasets on defining gene structure, we used SICILIAN[32] to uncover potential splice (exon-exon) junctions from sequence reads that mapped with high confidence to discontinuous positions along the genome (Fig. 2a-b, Table 2). For each discontinuous sequence read, SICILIAN computes a statistical confidence value based on features that influence sequence alignment (e.g., sequence entropy of the read, sequence mismatches, number of mapping locations for the read in the genome), and then read-level statistical values are aggregated to compute a final confidence score (empirical *p*-value) for each potential splice junction. The current mouse lemur genome annotation (NCBI Refseq Annotation Release 101 assembly Mmur 3.0, genome size ~2.5×10^9 bp) has 212,198 previously assigned splice junctions, 41% fewer than human (358,924 junctions in RefSeq hg38, assembly GRCh38.p13, genome size ~3.1×10^9 bp) and 33% fewer than mouse (319,497 junctions in RefSeq mm10, assembly GRCm39, genome size ~2.7×10^9 bp), suggesting that thousands of splice junctions remain to be discovered in lemur. Application of SICILIAN to both our 10x and SS2 scRNA-Seq data computationally supported nearly all (98%, 202,802 junctions) of the currently annotated splice junctions in NCBI's Refseq Annotation Release 101 for Mmur 3.0, accounting for 9.4% of identified junctions (Category A in Fig. 2b-d). Additionally, it uncovered 67,672 new junctions in which both 5' and 3' splice sites had been separately annotated (Category B) and 274,991 new junctions between an annotated and a novel splice site in the gene (Category C). These new junctions detected by SICILIAN were supported by an average of 3,711 (Category B) and 346 (Category C) unique reads mapped per junction across the entire dataset. SICILIAN also detected many new junctions that did not involve annotated splice sites (Category D) or map to known genes (Category E), but these were supported by fewer reads

(Category D, 98 reads/junction; Category E, 121 reads/junction) indicating some could result from natural "noise" in splicing[33,34] or are highly cell-type specific. Strikingly, among our detected junctions not currently annotated in NCBI's Refseq, 18,804 junctions are conserved and annotated in either human (RefSeq hg38) and/or mouse (RefSeq mm10) genomes, most of them (59%) belonging to category B (Fig. 2c-d). These results suggest that organism-wide scRNA-seq combined with rigorous statistical methods such as SICILIAN is a powerful means of deciphering RNA splicing and gene structure in a newly characterized genome.

To identify cell type-specific splicing events, we carried out a differential splicing analysis based on MANOVA, a statistical test of whether splice site usage for a gene, inferred from splice junction reads in scRNA-seq datasets, differs between cell types. MANOVA analysis identified 545 genes that are significantly differentially spliced across cell types in the same tissue for at least two individuals (Table 3). For example, myosin light chain *MYL6* is alternatively spliced to either include or exclude (skip) exon 6, and in most cell types both isoforms are produced. However, the ratio of the *MYL6* isoforms can differ dramatically among cell types (Fig. 2e, Fig. S2a). In pancreatic beta cells and alpha cells, most transcripts include exon 6, whereas in pancreatic ductal and acinar cells, as well as most immune and germ cells, almost all transcripts exclude it (Fig. 2e, Fig. S2a). *CAST*, a regulator of membrane fusion, also exhibits differential splicing of its exon 7, which is included in about 50% of transcripts in endothelial cell types but is almost always skipped in every other compartment (Fig. 2f, Fig. S2b). In addition, several genes including *MLF1, TPD52L2, FAM92A, CHPT1*, show sperm-specific splicing patterns and are differentially spliced across the stages of spermatogenesis (Fig. 2g, Fig. S2c-f).

3. Organism-wide scRNA-seq aids gene annotation

Gene identity assignments in newly sequenced genomes have traditionally relied on phylogenetic sequence comparisons and chromosomal position, with each assigned a name corresponding to that of the characterized homolog in other species with greatest sequence similarity and conserved gene order along the chromosome, indicating a direct evolutionary relationship (ortholog). However, such analyses can be ambiguous, returning no significant homologies or uncovering multiple homologs with similar levels of sequence identity and evolutionary relatedness, obscuring the true ortholog[35]. Hence, about a quarter of the genes (~7,600) in the current mouse lemur NCBI genome annotation (assembly Mmur 3.0[22], NCBI Refseq Annotation Release 101) have been assigned only a locus identifier (e.g. "Loc__" or “orf”) and no formal gene name/symbol or description (“uncharacterized genes”), and another quarter (~8,000) have been assigned a locus identifier and initial description based on sequence homology but no gene name/symbol (“unnamed genes”) (Fig. 3a). The fractions of these unnamed and uncharacterized genes in the current lemur genome annotation are much greater than those of human and mouse, including protein coding genes. We reasoned that organism-wide gene expression profiles could complement the classical approaches to gene orthology assignment and naming by identifying the sequence homologs with the most conserved expression pattern (the "expression homolog").

We used SAMap[36,37] to find the mouse and human homologs of each mouse lemur gene with the most similar patterns of cell type specific expression across 32 orthologous cell types in lung and muscle, carefully curated in the same way for all three species. This identified 1,279 expression homolog triads in lung and 1,686 in muscle (Table 4), most of which (91% and 89% in lung and muscle respectively) are triads of named orthologous genes across the three species,

substantiating their assignments as orthologs by traditional approaches (e.g., *ABCA3* is selectively expressed in lung AT2 cells from all three species) (Fig. 3b-d, Fig. S3). In addition, the conserved expression across the three species of 39 (3%) lung and 83 (5%) muscle orthologous gene triads where the lemur locus remains unnamed by NCBI (i.e., unnamed ortholog in Fig. 3b-c) strongly suggests that the lemur gene is the correct ortholog and should be named accordingly (e.g., lemur locus *LOC105861217* shares expression with its human and mouse ortholog *DYNLRB2* in lung ciliated cells) (Fig. 3e). This also includes instances where multiple unnamed lemur loci are assigned the same gene description by NCBI, ambiguity likely resulting from the traditional annotation pipelines relying principally on sequence homology. For example, both lemur loci *LOC105862649* and *LOC105862489* share the description "monocyte differentiation antigen CD14-like" in NCBI. By comparing their expression patterns across species, we were able to identify *LOC105862649* as the likely lemur *CD14*, which is expressed in lemur myeloid cells similar to human and mouse *CD14* (Fig. 3e). In contrast *LOC105862489* is sparsely expressed (<10 cells) in the atlas so may be a pseudogene. In other instances, we identified expression homolog triads with incomplete orthology assignments such as lemur locus *LOC105874770* (Fig. 3f), which shares expression in lung ciliated and fibroblast cells with the human ortholog *ALDH1A1* as well as with the mouse homolog *ALDH1A1*, though the mouse gene has not been assigned as an ortholog to the lemur gene by NCBI or Ensembl. The similar expression pattern between mouse and lemur provides strong evidence that these two genes are indeed orthologs currently missed by standard annotation pipelines.

Surprisingly, we also identified a small fraction of genes (71/1279 (6%) and 98/1686 (6%) in lung and muscle, respectively) whose expression patterns were not conserved with their assigned orthologs (Fig. 3c, g). For example, lemur *RAMP1* is highly expressed in lemur lung

endothelial cells, whereas its orthologs in human and mouse is only sparsely expressed in the lung (human *RAMP1*) or expressed in myeloid cells instead (mouse *RAMP1*). When compared to its homologs, we found that lemur *RAMP1* shares a lung expression pattern most similar to *RAMP2*, which is abundantly and selectively expressed in endothelial cells across all three species. This suggests that lung endothelial and myeloid cells have species-specific responses to hormones including CGRP and adrenomedullin through the RAMP/CALCRL receptor complexes[38–40]. Similarly, we identified such expression pattern changes for human *RAMP3* (top expression homolog being *RAMP2*), lemur *TEKT*3 (expression homolog *TEKT1*), mouse *Adamts15* (expression homolog *ADAMTS8*), and mouse *Cpxm1* (expression homolog *AEBP1*) (Fig. 3g). These exemplify rare instances where species-specific adaptations have dissociated gene expression patterns from their conserved sequence/structure. Examining expression homologs thus provides another dimension for orthology assignment, gene naming, and exploring evolutionary diversification in expression and function, which can be further enhanced by expanding the analysis to additional organs and species.

We used our organism-wide atlas to enhance gene annotation of the major histocompatibility complex (MHC), one of the most biomedically important vertebrate gene complexes because of its central role in the adaptive immune response, but also extremely difficult to assemble and annotate because of its evolutionary plasticity[41–45]. The MHC locus encodes arrays of ligands for T cell receptors (TCRs) and other families of lymphocyte receptor[46], with some MHC genes among the most polymorphic in the genome[47] due to point mutations and gene duplications/deletions; this genetic diversity individualizes immune systems and their response to infection, facilitating recognition of diverse pathogens and survival of individuals and entire species facing new and virulent pathogens.

MHC class II genes encode structurally-related alpha (A genes) and beta (B genes) subunits of transmembrane heterodimers expressed on dendritic and other professional antigen-presenting cells and loaded with exogenous peptides derived from phagocytosed pathogens and proteins, which serve as ligands for TCRs on helper T cells to initiate an adaptive immune response[48]. Our allele-specific analysis of expression of MHC class II genes across the full atlas shows that each individual mouse lemur expresses one or two alleles of each MHC class II gene except *DOA* and *DOB*, which were not expressed in any of them, supporting their designation as pseudogenes based on sequence analysis[49] (Fig. 4a, c). The other class II genes all appear to be single, non-duplicated genes, including two (*DQA, DQB)* that are currently annotated as duplicated (assembly: Mmur 3.0[22], annotation NCBI Refseq Annotation Release 101) and one (*DRB*) as triplicated, the latter from failed assembly of two genomic fragments of the second *DRB* allele[50] (Fig. S4b). *DQA, DQB, DRA,* and *DRB* are expressed at higher levels and in more cells than the other class II genes (*DMA, DMB, DPA, DPB*)[50] (Fig. 4d).

MHC class I genes encode the variable alpha chain that together with beta-2 microglobulin form widely-expressed transmembrane heterodimers that are loaded with peptides derived from cytosolic proteins, which serve as ligands for TCRs on cytotoxic T cells, or for KIRs (killer cell immunoglobulin-like receptors) or KLRs (killer cell lectin-like receptors) on NK cells, to destroy cells expressing viral, tumor or other "non-self" proteins[46,48]. No expression was detected in our atlas of any of the MHC class I genes in the MHC locus at chromosome 6 (Fig. 4a), supporting their assignment as pseudogenes from sequence analysis[51]. All of the expressed mouse lemur class I genes mapped to a separate, duplicated cluster on chromosome 20q (Fig. 4b), an organization unique to mouse lemur[50] (Fig. S4a). Our allele-specific analysis of the expression of these genes across the atlas identified 11 expressed genes in the cluster, four

(*Mimu-168, -W03, -W04, -249)* we designate as "classical" from their high and widespread expression, and the rest we designate “non-classical” (*Mimu-180ps, -191, -202, -208, -218, -229ps, -239ps)* including three previously thought to be pseudogenes (*Mimu-180ps, -229ps, -239ps*) based on sequence analysis (Fig. 4b-d). Three of the classical genes (*Mimu-168, -W03, -W04)* are very similar in sequence and appear to be divergent allelic lineages, with some individuals (L1, L3) missing (or not expressing) one or more of the three genes. Three of the non-classical genes (*Mimu -208, -218, -229ps*) are also missing (or not expressed) in two or three individuals (Fig. 4c), further highlighting the polymorphic nature of the gene complex. The current genome assembly (Mmur 3.0[22]) contains gaps separating the complex into three sequence islands, rearrangement of which better matches the available BAC sequence[51], as detailed in Guethlein et al.[50]; long range sequencing is needed to reveal the full number and order of genes for each individual (Fig. 4b).

4. Mapping immune gene expression across the body

The atlas can be used to map the expression of any gene or gene set across the body, for example to identify the most selective markers for each cell type (see accompanying Tabula Microcebus manuscript 1[6]) or to map hormone-producing and receiving cells to elucidate lemur endocrinology (see accompanying hormone atlas manuscript[40]). To explore the lemur immune system, we examined expression of key sets of immune genes. One set was the MHC genes, which as described above function in the adaptive immune system's recognition and response to pathogens and tumors and are also implicated in a wide variety of chronic and autoimmune diseases (e.g., Type 1 diabetes, inflammatory bowel disease, asthma). Their closely related and highly polymorphic sequences make MHC genes challenging to expression profile as well as annotate[52]. After resolving their identities (Section 3), we used the ascertained unique sequences

in each gene to discern the expression level of every MHC gene in every cell in the atlas (Fig. 4d).

The designated classical class I MHC genes were highly and broadly expressed (Fig. 4d), consistent with their general role in presentation of peptides derived from cytosolic proteins to *CD8+* T cells. However, expression of class I MHC genes varied substantially between compartments, with highest expression in the endothelial and immune compartments, intermediate levels in stromal and epithelial compartments, low in neural and almost undetectable in the germ compartment (Fig. 4d). Even within a compartment there were significant differences among cell types. For example, *CXCL10+* capillary cells and lung capillary aerocytes showed the highest class I gene expression in the entire atlas, and within the immune compartment expression was lowest in the erythroid-megakaryocyte lineage (Fig. 4d, Fig. S4c). Although class I gene expression was generally low in the neural compartment, non-myelinating Schwann cells were a striking exception: they showed high class I expression across multiple tissues (Fig. 4d, Fig. S4d), suggesting a special immune role protecting the peripheral nervous system. In the germ compartment, class I expression was detected at early stages of spermatogenesis but then rapidly declined[50] (Fig. 4d). Across the atlas, classical and non-classical class I genes were coordinately expressed, albeit the latter at much lower levels of expression (Fig. 4d, Fig. S4e). However, in sperm and some cortical and brainstem neurons, overall expression of class I genes was low and the non-classical genes were expressed at higher levels than the classical genes (Fig. S4e).

Class II MHC genes were more specifically expressed than class I genes (Fig. 4d), as expected from their role presenting fragments of extracellular pathogens to helper (*CD4+*) T lymphocytes to regulate their response to an infection, such as phagocyte recruitment and local

inflammation or activation of B cells to enhance their humoral (antibody) immune response. Class II MHC genes were expressed in "professional" antigen-presenting cells: dendritic cells, macrophages, and B cells but diminished in plasma cells (Fig. 4d). They were also expressed across the endothelial compartment of all tissues, similar to humans but not rodents[53]. They were expressed at especially high levels in several specialized capillary subtypes (Fig. 4d, S4d) including *CXCL10+*, *FAB5- RBP7-* cells, lung aerocytes, and kidney glomerular cells, as well as the ascending vasa recta of the kidney. Although there was little expression in the stromal compartment, two stem cell niche cells (adipo-CXCL12-abundant reticular (CAR), osteo-CAR cells, see below) and some mesothelial subtypes were interesting exceptions (Fig. 4d). There was also little expression in the epithelial and neural compartments, except for specific lung epithelial cell types (ciliated, AT2) and, in one individual (L2), the non-myelinating Schwann cells of several tissues with specific disease and inflammatory processes (Fig. S4d, Fig. 5b).

Mapping expression across the atlas of the identified immunoglobulin and B cell receptor genes (Table 5) established many classical features of B cell development and function in lemurs (Fig. 1i). Nearly all B and plasma cells expressed a dominant heavy chain and light chain isotype (see Methods). Most expressed an *IGG* or *IGM* heavy chain, but *IGA*-expressing cells were also found and indeed were the prominent isotype in small intestine and pancreas, similar to humans and mice[54]. We detected expression of most major variable domain genes (V_H) families, with broadest expression of genes related to human *IGHV1, 3* and *4*, as in humans and mice[28,55,56]. For Ig light chains, lambda was more commonly expressed than kappa, in contrast to human and mouse B cells where kappa predominates[57]. The most important region for antigen binding is the complementarity-determining region of the heavy chain (*CDRH3*), which is at the junction of the V, D and J genes created by V(D)J recombination; mapping the distribution of *CDRH3* length

across lemur Ig transcripts showed they were consistently shorter than in humans and more similar to the lengths in mice (Fig. S5)[58]. Cells expressed either membrane-bound or secreted forms of the Ig subtypes, and identification of 8 putative B cell lineage clones based on similarity of their *CDRH3* sequences (Fig. S5) provided evidence for heavy chain class switching during B cell development and clonal expansion and distribution of daughter cells to different organs.

We mapped expression of all chemokine receptors (24 genes) and their cognate ligands (34 genes) across the atlas (Fig. 5a, Fig. S6, Fig. S7, Table 6) to provide insight into the control of immune cell trafficking and interactions throughout the body. This identified subpopulations of bone marrow adipocytes ("adipo-CAR" cells) and osteoblast-like cells ("osteo-CAR") that express chemokine *CXCL12* and resemble the mouse adipogenic and osteogenic progenitors (see adipocyte section below)[59–61]. Expression of *CXCL12* by these progenitors is known to retain in the marrow hematopoietic progenitor cells expressing its receptor *CXCR4*, downregulation of which in maturing neutrophils may mediate their release into circulation (Fig. 5b)[62,63]. *CCL20* is selectively expressed in a variety of epithelial cell types (small intestine, colon, skin, tongue, and bladder) that could recruit various *CCR6*-expressing immune cell types, including dendritic cells to mucosal lymphoid tissue (Fig. S6)[64]. Likewise, *CX3CL1* selectively expressed by cortical and brainstem neurons could recruit or regulate brain macrophages (microglia), which specifically express its receptor *CX3CR1*[65,66]. We also identified signals that could traffic immune cells to lymph nodes. B cells, *CD4+* T cells, and mature dendritic cells (DCs) express receptor *CCR7*, and one of its ligands (*CCL21*) is selectively expressed on a subset of lymphatic endothelial cells (designated *CCL21+*) so could direct migration of these immune cells to and along afferent lymphatic vessels[67]. *CCL21* and *CCL19*, another *CCR7* ligand, are expressed in lymph node reticular cells so could attract arriving cells into the node[68]. This may be aided by *ACKR4*, a

decoy receptor for CCL19/CCL21, expressed in other lymphatic endothelial cells, epithelial cells of the skin (where it aids migration of Langerhans cells and DCs into lymph nodes[69]), as well as mesothelial cells and lung alveolar macrophages, and thus could enhance local gradients of CCL19 and CCL21 in and around lymph nodes[70]. Reticular cells also express *CXCL13* and *CXCL12*, which could help attract mature B cells that express the CXCL13 receptor (*CXCR5*) and other immune cells that express the CXCL12 receptor (*CXCR4*)[71,72]. The expression patterns of *RARRES2* and its receptor (*CMKLR1*) suggest they play a similar role for monocytes and macrophages in lymph node trafficking[73].

In addition to the stereotypical expression of chemokines and/or receptors across lemur individuals and tissues as in the above examples, which underlie normal homeostatic immune cell trafficking, we also uncovered examples of disease-associated expression of the chemokine pathways in specific individuals and tissues, consistent with the clinical data and histopathology[74]. These elucidate the lemur inflammatory response (Fig. 5b). For example, we found that the three "interferon gamma-induced chemokines" (*CXCL9, CXCL10, CXCL11*) were co-expressed at high levels in several diseased or inflamed tissues: alveolar and interstitial macrophages and capillary cells of L4's lung beset with pulmonary hemorrhage, and subpopulations of capillary and mesothelial cells in L2's inflamed mesenteric fat – presumably recruiting or retaining the *CXCR3*-expressing major dendritic cell types (conventional, plasmacytoid, mature) and other leukocytes infiltration into the inflamed tissue (and possibly suppressing angiogenesis)[75,76]. Similarly, a subset of neutrophils (designated as *CCL13+* and *IL18BP+* subtypes), monocytes, and macrophages express higher levels of the potent neutrophil chemoattractant *CXCL8* with its receptors *CXCR1* and *CXCR2* broadly expressed across all

neutrophils, a means of recruiting additional neutrophils and amplifying inflammation (see Section 5)[77].

5. Mapping immune cell development and activation across the body

Because our atlas includes bone and bone marrow, blood, spleen, and other organs, we could map the mouse lemur hematopoietic program from progenitors in bone marrow (see Fig. 3i in the accompanying Tabula Microcebus manuscript 1[6]) through their maturation, dispersal and differentiation throughout the organism, as well as their activation in disease. In the integrated UMAP (Fig. 6a, Fig. S8a-d), four major lineages radiate out from hematopoietic precursors, recapitulating the developmental trajectories of neutrophils; of monocytes, macrophages, and dendritic cells; of erythrocytes and platelets; and of proliferating lymphocytes.

Neutrophils are an abundant class of circulating leukocytes and among the first to arrive at an infection, where they ingest microbes and release granules containing enzymes that kill them. We used canonical neutrophil marker genes from other species (*CSF3R+*, *MSR1-*) to identify ~59,000 developing, proliferating, and mature neutrophils in the lemur atlas, and to establish that the trajectory emanating from hematopoietic precursors and granulocyte-monocyte progenitors recapitulated neutrophil development, allowing us to deduce its full developmental expression program (Fig. 6b). This showed sequential expression of granulopoiesis genes (Fig. 6c), which are specific to the type of granule produced at each stage of human neutrophil maturation. Neutrophils along the trajectory successively express markers of azurophilic (primary) granules (*AZU1*, *MPO*, *ELANE*) of myeloblasts and promyelocytes, specific (secondary) granules (*LTF*, *CAMP*, *LCN2*) of myelocytes and metamyelocytes, gelatinase granules (*MMP9*, *ARG1*) of band neutrophils, and finally the secretory vesicles (*ALPL*, *MME*) of mature segmented neutrophils[78,79] (Fig. 6c). Thus, lemur neutrophil granule biogenesis occurs in

a similar progression and uses similar genes as in human. This includes certain antimicrobial enzymes absent or lowly expressed in the corresponding mouse granules, such as defensins (*DEFA1*, *DEFA4*), bactericidal/permeability increasing protein (*BPI*), alkaline phosphatase (*ALPL*), and arginase (*ARG1*) (Fig. 6c)[80]. Lemur neutrophils also express multiple human neutrophil genes (*AZU1, CCL14, CCL16, CXCL8, IL32, TCN1, FCAR (LOC105877181), S100A12*) not present in the mouse genome (Fig. 6h, see also Section 7), indicating evolutionary elaboration of neutrophil function in primates[80–82].

Mapping the neutrophil maturation profile to different tissues and in different individuals provided insight into the development, activation, and function of neutrophils (Fig. 6d). Lemur L4 showed the expected distribution of neutrophils in health: earliest progenitors (likely myeloblasts, promyelocytes, myelocytes, metamyelocytes), and maturing neutrophils (likely band neutrophils) localized to bone and bone marrow except for a small number in spleen, lung, and hypothalamus/pituitary gland, whereas blood and all other tissues were enriched for mature neutrophils. No activated neutrophils were detected.

By contrast, lemur L2, who had an elevated white blood cell count (32.1 K/uL, normal range 4.5-11 K/uL in humans) dominated by neutrophils (differential: 91% neutrophils, 4% lymphocytes) and Klebsiella pneumonia (presumably secondary to lung metastases of uterine carcinoma, see Section 6) and cystitis, showed widespread neutrophil activation. This manifested as neutrophils at the end of the developmental trajectory, nearly all of which were from her infected organs (e.g., lung, kidney, bladder, Fig. 6d). These advanced neutrophils show downregulation of mature neutrophil markers (e.g., *SELL/L-selectin/CD62L, MMP9, CXCR1, ITGAL*) and induction of chemokines (*CXCR4, CXCL8/IL-8, CCL5/RANTES, CCL8 (LOC105885739)*, and *CCL4* (*LOC105881712*)), indicators of neutrophil activation that facilitate

extravasation (downregulation of L-selectin), homing to sites of inflammation (*CXCR4*), and recruitment of additional neutrophils (*CXCL8, CCL5*) (Fig. 6c)[77,83–86]. Some of these genes are also markers of neutrophil ageing and clearance (increased *CXCR4*, decreased *L-selectin* and *CD47* ("don't eat me" signal)), implying senescence and clearance rapidly follow activation[87,88] (Fig. 6c, Fig. 5b). Subclustering of L2's neutrophils in each tissue revealed two subpopulations of activated neutrophils that emerged from the main cluster of maturing/mature neutrophils and whose abundance differed by tissue (Fig. 6d, Fig. S9a). One activated neutrophil group, designated *CCL13+*, expressed chemoattractants for monocytes and eosinophils (*CCL15, CCL13* (*MCP-4/LOC105859268*), *CCL2-like* (*MCP-1/LOC105859340/LOC105885684*)) and was present in lung, the site of metastatic tumor and pneumonia, and sparse in kidney (Fig. S9b)[89–91]. The other neutrophil group, designated *IL18BP+*, showed enriched expression of genes including *IL18BP, ANXA5, SQSTM1, CD9, PTGR1,* and *CCR5*, many not previously implicated in neutrophil biology (Fig. S9b). This subpopulation was dominant in bladder (site of suppurative cystitis) and lung, and sparse in the kidney and perigonadal fat (Fig. S9b). These results suggest that local factors dictate the specific neutrophil activation pathway. We also uncovered a global response to neutrophil activation: the distribution of circulating neutrophils in L2 was shifted (compared to L4) toward more immature stages of the developmental trajectory, a molecular demonstration of the classical "left shift" seen in blood smears of human patients when immature neutrophils of bone marrow are released into circulation to replenish mature neutrophils recruited to an infection site (Fig. 6d)[92].

Unlike the widespread infection in L2, L3 did not have any detected infection, and although she too had uterine cancer (Section 6), her neutrophils showed the normal expected trajectory (similar to L4) except in the uterus where activated *IL18BP+* neutrophils dominated,

presumably in response to the tumor (Fig. 6d). Lemur L1 also showed a distinctive global neutrophil pattern, with neutrophils from across the developmental trajectory in circulation, presumably from dysregulation of granulopoiesis from severe fibrous osteodystrophy in his skeletal bones (Fig. 6d). His lung also harbored both *IL18BP+* and *CCL13+* activated neutrophils, perhaps from his emphysema, pulmonary edema, or suppurative rhinitis and tracheitis.

We also mapped development and trafficking across the body of the identified monocytes (*CSF1R+*) and macrophages (*CSF1R+*, and *C1QA+* or *MARCO+* or *MRC1+*), another lineage of phagocytic leukocytes that works with neutrophils to engulf and destroy microbial pathogens and foreign substances, as well as tumor cells and other abnormal cells and debris. Like neutrophils, monocytes formed a continuous and homogeneous trajectory emanating from hematopoietic precursor and granulocyte-monocyte progenitor, with the exception of disease-associated bladder and perigonadal fat monocytes in L2 as described below (Fig. 6a, e, Fig. S10e). Interestingly, we did not otherwise detect biologically meaningful subclusters of monocytes in any tissue, or identify subpopulations using canonical markers that distinguish human (classical, intermediate, and non-classical) or mouse (patrolling and pro-inflammatory) monocyte subtypes, suggesting lemur monocytes may be less specialized (Fig. S10d).

In contrast to neutrophils and monocytes, lemur macrophages were highly diverse: FIRM-based integration revealed many distinct, tissue-specific subpopulations (Fig. 6e). Many of the tissue-specific subpopulations clustered near monocytes, suggesting they arise from circulating monocytes that undergo tissue-specific maturation programs. Other subpopulations separated from the monocyte trajectory, suggesting these are stable populations of tissue-resident macrophages that colonized the tissue early in development. Indeed, many expressed genes

known to mark tissue-resident macrophages in human and mouse (Fig. 6f, Fig. S10a-b). CNS macrophages expressed microglial markers (*CX3CR1, P2RY12, SALL1*)[93,94]; lung macrophages that separated from monocytes and lung interstitial macrophages (*LOC105862649 [CD14-like] CD86*)[95] expressed alveolar macrophage markers (*SIGLEC1/CD169, MARCO, PPARG, LPL*)[96–98]; spleen macrophages expressed red-pulp macrophage markers (*SPIC, VCAM1*) as well as outer marginal zone (*LOC105885453 [SIGN-R1/CD209-like]*) and metallophilic macrophage (*SIGLEC1/CD169*) markers[99,100]; and liver macrophages expressed slightly higher levels of Kupffer cell marker *ID3*[101]. Liver macrophages further separated into two major subpopulations, one that clustered near monocytes and expressed canonical mouse macrophage marker *ADGRE1* (*F4/80*), and another (*MS4A7+*) that was *ADGRE1*-negative and clustered with a small intestine subpopulation, which also expressed *MS4A7* but showed no specific expression of resident intestinal macrophage markers (*LOC105869074 [CD163], TIMD4, IL10*)[102] (Fig. 6e, Fig. S10b). Recent profiling of human liver macrophages indicate they too can be separated into two subpopulations [103–105], although the differentially-expressed genes are distinct from those that distinguish the lemur subpopulations. We also identified another intestinal macrophage subpopulation (*MS4A7*-negative, *MMP1+, MMP9+, MMP10+, TNFRSF6B+, JAML+*) and a macrophage cluster (*MRC1+, CLEC10A+, GAS6+*) from five tissues (fat, bladder, tongue, kidney, limb muscle) whose profiles did not resemble that of any known resident macrophages (Fig. 6e, Fig. S10b). Bone macrophages specifically expressed osteoclast markers (*CTSK, ATP6V0D2* and *NFATC1*)[106] , although they did not cluster separately (Fig. 6e-f, Fig. S10b).

We identified a population of inflammatory monocytes in lemur L2's bladder and perigonadal fat, teaming up with the activated neutrophils in these tissues to battle the local infections (cystitis) and metastatic tumor. These monocytes clustered near macrophages from the

same tissues (Fig. 6e), and selectively expressed inflammation-associated genes, including *EREG, IL23A, AREG, CSF3, IL1A, TNFRSF18, SERPINB2, HES1, PTGES, CD274/PD-L1*, which were also expressed at lower-levels in L2's bladder macrophages, defining an inflammatory gene signature (Fig. S10c). In neither these nor other tissues did we distinguish M1 ("classically activated") vs M2 ("alternatively activated") macrophages based on commonly used markers[107,108] (Fig. S10d), though this distinction in other systems has blurred[109].

Lymphocytes showed no apparent developmental trajectory beyond proliferating lymphocyte progenitors (Fig. 6a, Fig. S8a-d). B cells formed two discrete clusters, mature B cells (*CD19+, MS4A1+*) and antibody-secreting plasma cells (*JCHAIN+, MZB1+, XBP1+*), with few evident intermediates (Fig. 6a, Fig. S8a-d). This suggests B cell development from proliferating progenitors and their differentiation into plasma cells are rapid with few standing intermediates, or that tissues containing such intermediates were not among those profiled. And, unlike macrophages, which exhibited many tissue-specific and some individual-specific populations, B cells appeared more molecularly homogenous across the atlas. Indeed, the only transcriptionally-distinct subpopulation was from pancreas (and nearby lymph nodes) and designated *SOX5+* B cells (*SOX5+, BHLHE41+, IL10RA+, CXCR6+, HCK+*) (Fig. S8e). Similarly, mature T cells, NK cells, NKT cells, and innate lymphoid cells formed a large, isolated cluster (Fig. 6a, Fig. S8a-d), belying their well established functional distinctions.

6. Cellular and molecular characterization of primate diseases and physiology

One reason for establishing mouse lemurs as a model organism is for investigations of primate-specific biology and disease. The mouse lemurs analyzed here were elderly and their clinical condition necessitated compassionate euthanasia; from the organism-wide expression profiling and full necropsies[74] a variety of human-like pathologies were identified. These

include: chronic kidney disease and metabolic bone disease (renal osteodystrophy), intracerebral hemorrhage/stroke, aortic atherosclerosis and coronary artery arteriosclerosis, cardiomyopathy, bacterial pneumonia, thyroid papillary adenoma, osteoarthritis, cataracts, periodontitis, and otitis media and externa.

Both females (L2, L3) had endometrial cancer (Fig. 7a-e), the most common malignancy of the female reproductive tract and fourth most common cancer in U.S. women (3.1% lifetime risk[110] (https://seer.cancer.gov/statfacts/), with increasing incidence and mortality over the past decade attributed to an aging population and increasing obesity[111,112]. Type 1 endometrial carcinomas are of low-grade endometrioid histology and typically estrogen-induced with a favorable prognosis under standard therapy of resection and local radiation, whereas type 2 tumors are of high-grade endometrioid, or of serous or clear cell histology, commonly estrogen-independent and lack durable therapies. Animal models are limited: mice do not naturally acquire endometrial cancer, and although rats do, they and the engineered mouse models generally resemble type 1 rather than the high-grade, intractable type 2 tumors[113]. The spontaneous mouse lemur tumors were uncovered in the first female analyzed (L2) as a novel lung epithelial cell type (Fig. 7f) that curiously expressed high levels of oxytocin receptor (*OXTR*) (Fig. S11), a gene known to be highly expressed in female reproductive tissues and brain[114,115]. Comparison to cells across the atlas showed their expression profile resembled uterine cells of the other female analyzed (L3) (Fig. 7g-i, Fig. S11); necropsies established the diagnosis of primary uterine endometrial carcinoma with lung metastases (L2) or with local spread to mesenteric lymph nodes forming an abdominal mass (L3)[74]. Organism-wide atlases thus enable a new approach for identifying the primary site (and metastatic spread) of cancers of unknown origin, which comprise ~2% of all human cancers[116].

The presumptive primary tumor cells in the uterus, based on co-expression of the classical human endometrial and ovarian cancer serum serum markers Cancer Antigen 125 (*CA125/MUC16*) and *HE4/WFDC2*[117–119], showed enriched expression (relative to other uterine epithelial cells) of *OXTR*, as well as expression of *MYC* and *ERBB2*, two genes commonly amplified, overexpressed, or mutated in human type 2 endometrial tumors[120,121]. In addition, tumor expression of *inhibin-βB* (*INHBB*) likely produced INHBB homodimer (activin B), which promotes migration and invasion of human endometrial cancer cells and is correlated with higher grade tumors[122,123]. The lung metastasis also expressed *ERBB2*, its binding partner *EGFR* and highly specifically the ligand *EGF*, suggesting progression to autocrine mitogenic signaling during metastasis, and it lacked expression of estrogen receptor (*ESR1*) (Fig. 7j), loss of which also correlates with more advanced stage and grade human tumors[124]. Endometrial cancer of these two mouse lemurs thus molecularly, histologically, and clinically models the more aggressive form of the human disease; it can be used to explore susceptibility factors (e.g., genes, pregnancies/parity, obesity, age), pathogenetic mechanisms, and therapies including promising anti-angiogenic therapies targeting VEGF[125] and endocrine therapies[40,126] targeting ESR1[127] since both targets were expressed in the uterine tumors (though *ESR1* expression was lost in the metastasis[40]) (Fig. 7j). Human therapies might help control the lemur disease, which has long afflicted laboratory colonies[128].

One notable aspect of mouse lemur physiology is their dramatic seasonal/annual increase in body weight and daily change/variation in body temperature and metabolism as they enter a hibernation-like (torpor) state during the dry, resource-poor winter season, providing a valuable model for primate seasonal (circannual) rhythms and regulation of metabolism, adipose biology and body weight[11,15,129–131]. In our scRNA-seq profiles of four major fat depots (subcutaneous

adipose tissue (SCAT), mesenteric (MAT), interscapular brown (BAT), and perigonadal (GAT)), we identified adipocytes (Fig. 7k-l, Fig. S12a-c) and dozens of associated cell types including fibroblasts, pericytes, tendon cells, smooth muscle cells, reticular cells, mesothelial cells, arteries, veins, capillary cells, lymphatics, and all major types of immune cells. These fat depot adipocytes expressed many canonical mouse and human adipocyte markers including lipid biosynthetic and metabolic genes (e.g., *PNPLA2, FASN, GPD1, CIDEC*) and adipokines (*ADIPOQ, CFD*) (Fig. 7n)[132,133]. With these markers we also identified rare adipocytes in seven other tissues (aorta, bladder, bone, kidney, limb muscle, pancreas, thymus) (Fig. 7m, Fig. S12a, d). In total, we captured 1,231 adipocytes and adipo-CAR cells (1,222 by 10x, 9 by SS2) from two individuals (L2 and L4), most of which (>95%) express mature adipocyte markers (*CEBPA, GPD1, PPARG*) (Fig. 7n)[132].

Mouse lemur adipocytes showed two surprising features. Although they express most well-established adipocyte markers including those noted above and also *CIDEA, ADIG,* and *RBP4*, they curiously did not robustly express the classical adipocyte hormone leptin (*LEP*), which is highly expressed by human and mouse adipocytes and regulates food intake, energy expenditure, and body weight[134] (Fig. S12f). *LEP* expression was detected in only 0.6% of mouse lemur adipocytes (7 of 1,231 cells) and at a low level (average of 3.2 transcripts per 1e4 reads), and also sparsely and at even lower levels in several unrelated cell types (Fig. S12f). However, its receptor *LEPR* is selectively and highly expressed in a similar cellular pattern as in humans and rodents, including fibroblasts, tendon cells, endothelial cells, osteo-CAR cells and adipo-CAR cells (Fig. S12f). Perhaps *LEP* expression is inducible in lemur adipocytes depending on age, season, diet, or body weight[135], or some occult cellular source (or another gene) has usurped its function.

The other surprising aspect of mouse lemur adipocytes is the blurring of the distinction between white and brown adipocytes. Aside from the bone adipocytes ("adipo-CAR cells") that may be adipogenic progenitors (Fig. 7n, Fig. 5b), the bulk of the adipocytes formed two continuous populations (Fig. 7k-l, Fig. S12a-c). These populations were distinguished by expression of Uncoupling Protein 1 (*UCP1*), the canonical marker used in humans and rodents to distinguish the thermogenic brown adipocytes from the white adipocytes that do not express this gene[136]. We designate the *UCP1*-high population in mouse lemur as "brown-like adipocytes" because they also express higher levels of known thermogenesis regulators including genes involved in fatty acid metabolism (e.g., *CPT1B*, *SLC27A2*, *FABP3*), mitochondrial respiration (*LOC105876884* (*COX7A1*)), transport ATPase *LOC105862687* (*ATP1A2*), and adrenergic signaling mediated lipolysis (*KCNK3*)[137] (Fig. 7n). We designate the *UCP1*-low population (which includes a subpopulation distinguished only by total detected genes and UMI levels per cell but no biologically significant differentially-expressed genes) as "white-like adipocytes" because they are enriched in expression of many white adipocyte genes including *APOE*, *NNAT*, and *DPT*, despite their expression (albeit low) of the brown-defining gene *UCP1*[138–140]. Further blurring the white-brown distinction, both the classical brown marker *CIDEA* and the classical white adipokine *RBP4* were equally expressed across all adipocytes[141,142] (Fig. 7n). These seemingly mixed molecular signatures indicate that the distinction between white and brown adipocytes is less strong in *M. murinus*, and the observed continuum between the populations suggests potential interconversion between white-specific lipid storage and brown-specific thermogenesis properties, affording functional plasticity to accommodate the dynamic energy needs of this seasonally-cycling primate.

There was no exclusively brown-like fat depot among the depots surveyed: each site contained exclusively white-like adipocytes (GAT, SCAT, aorta, bladder, bone, limb muscle) or a mix of both (BAT, MAT, pancreas, kidney) (Fig. 7m). And, the different depots did not cluster separately and did not differentially express any biologically significant genes, except for gonadal adipocytes, which express higher levels of *S100A8/9/12*, *IL1B* and metallothionein genes *LOC105866476* (*MT2A*) and *LOC105866554* (*MT1E*) (Fig. S12e), all known in human and rodents to be expressed by adipocytes at fluctuating levels depending on inflammatory and feeding status and to correlate with insulin-resistance[143–146]. It will be critical to systematically study the seasonal changes in adipocyte gene expression in each depot to explore adipocyte plasticity, how it is controlled, and its role in seasonal physiology and behavior.

7. Expression patterns of primate-specific genes in mouse lemur

Mouse lemurs could be especially valuable for the study of primate-specific genes that are missing, or expressed differently, in other model organisms such as mouse. By comparing lists of orthologous protein-coding genes in human, lemur, and mouse (Table 7, compiled from NCBI, ENSEMBL and MGI homology assignments, see Methods), we identified 546 human genes for which there are orthologues in lemur (429 orthologs annotated in NCBI) but not mouse ("human-lemur" or "HL" genes, Fig. 8a, Table 8). Twenty-five of these are known to cause human disease or phenotypes (OMIM), including *EYS*, *RAX2*, and *ARMS2* associated with eye disease, *ANOS1* with Kallman syndrome, and *KRT85* with ectodermal dysplasia. Other HL genes play important roles in human physiology, such as motilin (*MLN*) and its receptor (*MLNR*) in gastrointestinal motility[147], surface proteins *CD1B* and *CD58* in antigen presentation to T cells; and *FCAR* (IgA receptor), *CXCL8*, *IL32*, and *S100A12* in the inflammatory response to infection. Gene set enrichment analysis shows that HL genes are enriched in transcription factor activity

and regulation (GO term: 000370, 0000977) and herpes simplex virus 1 infection (KEGG: 05168), including many that encode zinc finger proteins (Table 9). Nearly all (94%) of the 429 HL genes annotated by NCBI were expressed in the lemur atlas; their global expression patterns are shown in Fig. S13 (Table 8). Some were selectively expressed (or depleted) in specific compartments (168 genes) and/or in specific organs (97 genes) (Fig. 8b-d). Many of the selectively-expressed HL genes were specific to developing or mature male germ cells (e.g. *LUZP4, PASD1, LIPJ, IQCF2, TSPAN16*), immune cells (e.g. *OPRPN* in hematopoietic precursors, *FCRL2* in B cells and plasma cells, *C27H1orf186* in innate lymphoid cells, *AZU1* and *TCN1* in neutrophils, *C4BPB* and *CCL14* in monocytes and macrophages, *AHSP* in erythrocytes, and *ICAM3* and *CXCL8/IL8*, and *P2RY8* more broadly) and neurons of the cortex, brainstem, pituitary gland, and retina (e.g. *BEX5, DIRAS3, NAT16, PIWIL3, ZNF385D*), indicating substantial evolutionary gene plasticity in the germline, immune, and neural compartments (Fig. 8b-c, e, Fig. S13). But there were also examples of cell type-selective expression of HL genes in most other organs including small intestine and colon (*LGALS9* in enterocytes, *BEST4* in M cells), liver (*SERPINA4* in hepatocytes), kidney (*MLN, ENOSF1, SLC2A11* in tubule cells), bone and trachea (*C1QTNF8* and *FGFBP2* in chondrocytes), retina (*GUCA1C, GRK7*, and *EYS* in cones, *CLUL1* and *TMEM244* in rods) and uterus (*ADA2, MX2*) (Fig. 8d, Fig. S13).

DISCUSSION

Although the recent explosion of new model organisms has been fueled by the creation of reference genome sequences, and by the promise of CRISPR-Cas9 systems for targeted mutagenesis, what has remained limiting is the rich cellular and molecular understanding of a new model organism that typically takes decades to acquire. Our results show how a comprehensive molecular cell atlas of the mouse lemur[6] can be readily exploited in at least four ways to elucidate its genes, physiology, and diseases, and provide insights into primate biology and evolution.

First, we showed that the massive transcriptomic datasets uncover thousands of previously unidentified genes (Section 1) and hundreds of thousands of new splice sites (Section 2) that globally define lemur gene structures, splice isoforms, and their cellular expression patterns. Many are expressed only in specific cell types, so would be obscured in conventional bulk tissue profiling by more broadly expressed genes and isoforms. Their expression patterns also aid in gene naming and orthology assignments, by identifying the lemur homolog with the most similar expression pattern ("expression-log") to characterized genes in other species (Section 3).

Second, we showed that global and dynamic aspects of development and physiology can be revealed by comparing cell types and their gene expression patterns across the atlas. This was especially valuable for broadly distributed cell types such as immune cells (Section 5), and for gene products like hormones that function across the organism[40]. For example, we defined the full expression program of lemur neutrophils, from their birth from progenitors and initial maturation in bone marrow, to their release into circulation and dispersal and differentiation

throughout the organism, and finally their activation and turnover at sites of inflammation. We also identified chemokine signals and their cellular sources that likely regulate these key transitions (Section 4).

Third, by comparing cell profiles from different individuals we found individual-specific molecular cell types and states, many of which are disease or inflammatory states. A prominent example was the metastatic endometrial cancer in lemur L2, initially identified as a novel cell type in her lung. Its expression profile closely matched that of uterine epithelial cells, establishing a new use for cell profiling and organism-wide atlases in uncovering the primary site of cancers of unknown origin. And, in conjunction with her clinical data and full histopathology collected on autopsy[74], her molecular cell profiles provided unprecedented cellular and molecular insight into the full pathogenic sequence of the disease (Fig. 5b): from tumor initiation, its metastasis to lung, secondary infection leading to Klebsiella pneumonia, and the immune response in each affected tissue including recruitment of neutrophils, monocytes, macrophages and lymphocytes to the sites of the tumor and infection and to her inflamed bladder and adipose tissue, followed by trafficking of the inflammatory cells to lymph nodes, and the hematopoietic feedback response in bone marrow. As cell profiling becomes easier and cheaper, such "molecular cell autopsies" will provide an important new avenue for identifying diseases and the underlying pathogenic mechanisms and global responses in new model organisms like mouse lemur, as well as for new "systems-level" understanding of human diseases.

Finally, our analysis provides insight into primate-specific biology and evolution. This includes the expression patterns of the hundreds of genes present in human and lemur but missing in mouse, many of which are selectively expressed in the germline, immune, or neural

compartment (Section 7), plus the many genes present in all three species but for which their expression in specific processes such as neutrophil differentiation (Section 5) or spermatogenesis (see accompanying Tabula Microcebus manuscript 1[6]) is not conserved for the mouse ortholog. It also includes the aggressive, metastatic endometrial cancer that resembles the deadly human form of the disease, and the unusual molecular features of mouse lemur adipocytes such as lack of (or exceedingly low) expression of the classical adipocyte hormone leptin and the blurring of the molecular distinction between white and brown adipocytes, which could underlie the dramatic seasonal rhythms of mouse lemur and the associated changes in body weight and metabolism (Section 6). The atlas establishes a foundation for identifying and exploring such molecular and cellular aspects of primate-specific biology and diseases, as well as for the fascinating lemur-specific and human-specific evolutionary elaborations.

The approach developed here can be readily applied to the many emerging model organisms to establish molecular and cellular foundations for exploring their unique biology. Unlike the original mouse cell atlas obtained from cell profiles of genetically and environmentally homogenous animals[148,149], atlases for emerging model organisms will generally be opportunistic like this one and our recent human cell atlas[150]. Hence it will be critical to profile at least several individuals of the species, and to provide along with the molecular cell atlas extensive metadata for each individual, such as its distinguishing features, life history, and all clinical data including full autopsies and histopathology. This should then create both a working reference ("healthy") cell atlas — crucial for understanding normal physiology and for evolutionary comparisons — but also identify interesting variants like the mouse lemur cancer and inflammatory states described here. In this regard, we have demonstrated the value of

profiling all tissues from an individual, especially for elucidating cell types and the many physiological and pathological processes that occur or have impacts throughout the body. There is also great value in using the same profiling strategies and analytical approaches for each organ and organism to ensure fidelity of evolutionary comparisons. This becomes increasingly important as organism-wide evolutionary analyses expand beyond the three species compared here (lemur, human, mouse) to other emerging model organisms and ultimately the full tree of life.

ACKNOWLEDGMENTS

This work was supported by funding from The Chan-Zuckerberg Biohub, Howard Hughes Medical Institute and Vera Moulton Wall Center for Pulmonary Vascular Disease; unrestricted grant from Research to Prevent Blindness and NEI P30-EY026877 to the Stanford Department of Ophthalmology to Albert Y. Wu; The Hong Kong University of Science and Technology (start-up grant R9364), The Hong Kong University of Science and Technology Big Data for Bio Intelligence Laboratory (BDBI), and The Chau Hoi Shuen Foundation (R9056) to Angela Wu; NSF-DBI-1701984 to Anne D. Yoder; funding from Developmental and Stem Cell Biology Graduate Program to Aris Taychameekiatchai; funding from Hong Kong Research Grant Council (16307818, 16301419, 16308120, 16307221, C6021-19E), the Hong Kong University of Science and Technology (startup grant R9405), and The Hong Kong University of Science and Technology Big Data for Bio Intelligence Laboratory (BDBI) to Can Yang; funding from Shanghai Sailing Program (21YF1410600) to Jingsi Ming; NIA 1K99AG066963 to Thomas Ambrosi; NIH DP2AI138242 to Iwijn De Vlaminck; National Sciences and Engineering Research Council of Canada fellowship PGS-D2 to Michael F.Z. Wang; NIH R35 GM139517, R01 GM116847, R35 GM139517, NSF MCB1552196 to Julia Salzman; NSF Graduate Research Fellowship DGE-1656518 and a Stanford Graduate Fellowship to Julia Olivieri; Cancer Systems Biology Scholars Fellowship (Grant R25 CA180993) and Clinical Data Science Fellowship (Grant T15 LM7033-36) to Roozbeh Dehghannasiri; Urology Care Foundation Research Scholar Award Program and AUA Western Section Research Scholar Fund II to Hosu Sin; Stanford Graduate Fellowship/HHMI/NIH CMB Training Grant to Yue Zhang; American Cancer Society Postdoctoral Fellowship to SoRi Jang; Walter V. and Idun Berry Postdoctoral Fellowship to Andrea R. Yung; NSF Graduate Research Fellowship and Stanford Graduate Fellowship to

Youcef Ouadah; the Wu Tsai Neurosciences Institute Interdisciplinary Scholar Award to Shixuan Liu; European Community's 7th Framework Programme (FP7/2007-2013) under grant agreement Nbr 278486 (DEVELAGE), from the Fonds Unique Interministériel and Région Languedoc-Roussillon under grant agreement Nbr 110284 (DiaTrAl), and from the Fondation Plan Alzheimer (PRADNET) to Jean-Michel Verdier and Corinne Lautier; NIH R01 AI024258 to Peter Parham and Lisbeth A. Guethlein; NIH R01DC016892 to Wan-Jin Lu; NIH P30DK116074 to Yan Hang; NSF Graduate Research Fellowship to Connor V. Duffy; Postdoctoral Fellowships from the DFG (NE 2006/1-1) and California TRDRP (25FT-0011) to Patrick Neuhöfer; funding from Independent Research Fund Denmark (DFF–5053-00195) and Lundbeck Foundation (R232-2016-2459) to Jean Faru; NIH AG068667, AR073248 and AG036695 to Thomas A. Rando; funding from Wu Tsai Neurosciences Institute to Tony Wyss-Coray; NSF BCS 0647402 to Liza Shapiro and E. Christopher Kirk.

METHODS

Data availability

Tabula Microcebus data (including cell coordinates of FIRM-integrated UMAPs and MHC gene expression data) and cross-species datasets are in Figshare (https://figshare.com/projects/Tabula_Microcebus/112227). Tabula Microcebus web portal (https://tabula-microcebus.ds.czbiohub.org/) also allows for interactive visualization of the scRNA-seq data with cellxgene and the histological images of the mouse lemur tissues analyzed.

uTAR analysis to identify unannotated genes

To uncover unannotated transcriptionally active regions (uTARs), we employed the workflow developed by Wang et al.[24] for scRNA-seq data that identifies transcriptionally active regions (TARs), genome regions with abundant transcript alignments, using a previously published tool groHMM[151]. Briefly, all mouse lemur 10x datasets were aligned to genome assembly Mmur 3.0 using STAR with default parameters, without gene annotation indexing, and transcribed regions were predicted using the groHMM tool. TARs within 500 bps of another were combined into a single TAR and kept if it was expressed in at least 2 cells of the droplet-based atlas. The detected TARs were then separated into annotated TARs (aTARs) and unannotated TARs (uTARs), based on whether the region is currently annotated as a gene in NCBI Annotation Release 101 of Mmur 3.0. This identified that in the mouse lemur genome (2487 Mbps), aTARs and uTARs cover 284 and 42 Mbps, respectively.

To filter out transcriptional and sequencing noise from biologically significant uTARs, we then examined if a uTAR is differentially-expressed across cell types by Wilcoxon rank-sum test. This analysis was performed separately for each tissue and individual lemur. A differentially-expressed uTAR (DE-uTAR) is defined as a uTAR with a significant Bonferroni-

corrected *p*-value <0.05 from the Wilcoxon rank-sum test, expressed in ≥25% cells of a cell type (cells of the same free annotation), and cell type average expression level is ≥ 1.65 ($e^{0.5}$) times the average of other free annotation cell types in the same tissue. Some of the uTARs passed the differential expression test in multiple cell types and/or tissues. Together, a total of 4003 unique DE-uTAR were identified. To infer their identity, we applied BLASTn on each of the DE-uTAR against the nucleotide collection (nt) database (with a threshold of maximum e-value of 0.01 and a minimum bit score of 50), using either the entire length of the uTAR or the peak coverage region (full width at half maximum region around the absolute peak in coverage after Gaussian smoothing within the uTAR location). Occasionally, multiple uTARs aligned to the same gene in another species. The genome location and inferred homology of all DE-uTARs as well as their expression levels across the atlas 10x cells are provided in Table 1. There were 30 DE-uTARs with a BLASTn result that corresponded to one of the 2055 human genes without a mouse lemur ortholog, which are likely genes missed from NCBI's and Ensembl's annotations of Mmur 3.0 (Table 1). Note, 14 out of these 30 genes have no mouse ortholog, suggesting these are primate (human-lemur)-specific.

DE-uTARs were also classified as protein-coding or noncoding using Orpheum (Botvinnik et al 2021), with sequences containing 95% or more k-mers (k=9) matching a reference database of mammalian proteins from UniProt assigned as putatively protein-coding, and otherwise, assigned as noncoding. Coding sequences were then annotated using DIAMOND blastp[152] and noncoding sequences were annotated using Infernal cmscan (https://www.ebi.ac.uk/Tools/rna/infernal_cmscan/). Some uTARs had both coding reads and non-coding reads, potentially representing incompletely spliced transcripts or untranslated regions.

To detect the developmental trajectory of the sperm lineage cells in the testes 10x dataset by the uTAR expression data, we followed the same procedure as detecting the trajectory by gene expression data, which is described in the accompanying manuscript[6]. The analysis included a total of ~50,000 uTARs that have transcript reads in the testis dataset. Each uTAR was treated as a gene and data were first normalized for scRNA-seq library size (to 10,000 total uTAR transcripts per cell) and natural log-transformed. The top highly-variable uTARs (~1500) were then used for principal component analysis. The top 20 principal components that were not driven by extreme outlier data or immediate early genes were used to construct a 2D UMAP using cell-cell Euclidean distances as input. The pseudotime developmental trajectory was then identified as the density ridge of the data in this 2D UMAP by automated image processing. Cells were assigned to the trajectory by the shortest connecting distance. The cells' pseudo-time trajectory coordinates were linearly normalized such that the trajectory started at 0 and ended at 1, and then were compared to the pseudo-time coordinates derived from the gene-based trajectory using Pearson's correlation. The uTAR expression data were similarly pre-processed (normalized, scaled, and UMAP embedded) in other analyses, including when comparing the UMAP cell distribution patterns of the 10x colon dataset and in the silhouette coefficient analysis (Fig. S1a-b).

To measure the consistency of cell distribution patterns for 10x datasets embedded in a UMAP using the gene, aTAR, and uTAR-expression space, silhouette coefficient values were calculated for each dataset (separated by tissue and lemur individual, and sequencing channel). Cells were grouped according to the expert cell type designation (free annotation). The silhouette coefficient value for each cell *i* in a dataset was calculated as $s(i) = (b(i) - a(i))/ max\{b(i), a(i)\}$, where $a(i)$ is the mean in-group distance (mean distance of cell *i* to the other

cells in the same cell type) and $b(i)$ is the minimal out-group distance (minimal distance of cell i to any cell in the tissue of a different cell type). The cell silhouette coefficient values were then averaged across each dataset to derive the dataset-averaged silhouette value - an overall score representing how well each cell type co-clusters and separates from other cell types in the UMAP embedded space, with higher positive values representing better separation. The silhouette coefficient values were calculated separately using the cell-to-cell distances in UMAPs based on the gene expression, aTAR expression, and uTAR expression spaces, and then compared with a box plot (Fig. S1a).

The list of genes for each category in Fig. S1e were obtained as follows. Lists of the top n variable NCBI-annotated genes were derived by applying variance-stabilizing transformation to the entire 10x dataset and selecting the genes with top n transformed variance. The list of the genes annotated by Ensembl only (not by NCBI) were detected by comparing the genomic positions of mouse lemur genes annotated in the two databases, searching for the Ensembl-annotated gene with no overlap in NCBI. The primate (human-lemur)-specific genes are derived as in Methods - Analysis of primate-specific genes (Table 8).

To determine the amount of transcriptomic sequence information provided by the mouse lemur scRNA-seq atlas datasets, we estimated the total number of sequenced base pairs (bp) that mapped to the mouse lemur genome. For the 10x datasets, the total number of aligned paired-end reads (~1.63×10^{10}) was multiplied by the number of bp per read (90), equating to 1.46×10^{12} bp, though this number is inflated by PCR duplicates. However, summing the number of unique molecular identifiers (UMIs) across all cells in the atlas (~1.22×10^{9}) and multiplying it by the number of bp per read (90), equated to 1.10×10^{11} bp – 10-fold less than the estimate with total reads. For the SS2 datasets, the total number of aligned paired-end reads (1.11×10^{10}) was

multiplied by the number of bp per read (100), equating to 1.11×10^{12} bp (unique reads are not possible to identify in SS2 datasets). In comparison, the total number of bulk RNAseq bp used by NCBI for gene prediction/annotation of Mmur 3.0 equates to ~3.4×10^{11} bp (across ~3.4×10^{9} total reads), which was calculated by summing the number of sequenced bp for each RNAseq run (all tissues and generic samples included) – information obtained from the Sequence Read Archive (Biosample ID numbers and corresponding SRA link available at https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Microcebus_murinus/101/). However, this estimate is inflated because it includes bp from unaligned reads and does not correct for PCR duplicates.

SICILIAN splicing analysis

To identify splice junctions of mouse lemur transcripts, we employed SICILIAN, a statistical method for unbiased and annotation-free detection of splice junctions[32,153]. Raw sequencing reads from the atlas (both 10x and SS2 datasets) were first aligned to the mouse lemur genome assembly Mmur 3.0 by the STAR algorithm with parameters chimSegmentMin = 12, chimJunctionOverhangMin = 10, chimOutType = "WithinBAM SoftClip Junctions", and default values for the rest of the parameters. SICILIAN first extracts spliced reads that mapped to discontinuous regions of the genome. These reads could provide evidence for candidate splice junction sites but may also reflect sequencing noise or alignment error. For each of these potentially-spliced scRNA-seq reads, SICILIAN estimates a read-level confidence score, quantifying the probability that the alignment of the read is true. It then incorporates all the read-level scores for the reads aligned to each potential splice junction to compute an empirical *p*-value for the junction. The empirical *p*-value is computed for each SS2 cell or 10x channel within the dataset, then the median is calculated for these empirical *p*-values across the dataset.

Junctions with median empirical *p*-value <0.1 were deemed statistically significant and were selected for follow-up analysis. The threshold 0.1 was determined as the optimal point in the receiver operating characteristic (ROC) curve based on simulated data. We performed SICILIAN analysis separately on the reads from each lemur individual and scRNA-seq method (10x and SS2).

The detected splice junctions were compared to the junctions in transcripts annotated in the NCBI's Annotation Release 101 of mouse lemur genome assembly Mmur 3.0. The detected splice junctions were categorized into five types (see Fig. 2b). Type A refers to junctions that matched an annotated splicing pattern. Type B-D refers to junctions that aligned to an annotated gene but the specific splicing pattern is unannotated; type B contains junctions in which both the donor (5') and acceptor (3') splice sites are annotated but not previously paired (e.g., unannotated exon skipping); type C contains junctions in which one site is annotated and the other is not (e.g., annotated donor site but unannotated acceptor site); type D contains junctions in which both splices sites are unannotated. Type E refers to detected junctions that do not align to any annotated gene.

To examine if the detected splice junctions are conserved in human or mouse genomes, we used the UCSC LiftOver tool from the UCSC genome browser[154] and computed the fraction of junctions annotated in mouse and/or human by considering a junction as conserved only if it had successful LiftOver conversion to the other genome (i.e., both 5' and 3' splice sites in the lemur genome mapped successfully to unique coordinates in the other genome).

To identify cell-type specific splicing events, we performed MANOVA analysis separately on the 10x data from each tissue. Cell types with fewer than 10 cells or junctions present in fewer than two cells were removed from the analysis. Analyzed genes were limited to

ones that were expressed in all cell types of the tissue with at least two spliced reads mapping to the gene in each cell type. Let C be the set of cells, and J be the set of junctions for this gene. Consider cell $m \in C$ and junction $i \in J$ for a particular gene. Let $n_m^{(i)}$ be the number of reads mapping to junction i in cell m. Then define the fraction of junctional reads mapping to junction i in cell m as: $f_m^{(i)} = n_m^{(i)} / \sum_{j \in J} n_m^{(j)}$. Then calculate the dataset average fraction of junction i as: $f^{(i)} = \sum_{c \in C} n_c^{(i)} / \sum_{c \in C} \sum_{j \in J} n_c^{(j)}\, n_c^{(i)}$. Then let the scaled z score for junction i in cell m be defined as: $z_m^{(i)} = \left(\sqrt{\sum_{j \in J} n_m^{(j)}}\right)\left(f_m^{(i)} - f^{(i)}\right) / \sqrt{f^{(i)}(1 - f^{(i)})}$. MANOVA analysis was then performed using the cellular z scores of each junction for the gene as input and the cell type (free annotation) as output. This analysis generates, for each gene, a *p*-value of all cell types in the tissue having the same multivariate mean junctional expression. Benjamini-Hochberg correction was then applied to all *p*-values. Junctions with a corrected *p*-value less than 10e-16 were considered significantly differentially sliced across cell types.

SAMap analysis to study conservation of gene expression patterns across species

To compare expression patterns of homologous genes across human, lemur, and mouse, we used the SAMap method developed by Tarashansky et al.[36,37]. In addition to the lemur 10x cell atlas datasets, mouse and human 10x scRNA-seq datasets were retrieved from Tabula Muris Senis[148] and Tabula Sapiens[150], respectively. We applied SAMap to these datasets to compare the cell type expression patterns of homologous genes across the three species. While the SAMap algorithm does not require cell type labels, having comparable annotations simplifies the interpretation of the mapping results. Therefore, we limited the analysis to the tissues (lung and limb muscle) that were reannotated using the same standards as described in the accompanying

Tabula Microcebus manuscript 1[6]. This cross-species data with new unified cell type designation can be found on Figshare https://figshare.com/projects/Tabula_Microcebus/112227.

SAMap was used to simultaneously map genes and cells across the three species (human, lemur, and mouse). For each pairwise combination of species, SAMap first detects homologous genes by bi-directional BLAST of the transcriptomes of the two species, as annotated by Ensembl and NCBI. A cross-species gene-to-gene graph is then generated with edges connecting a gene in one species and a homologous gene in the other species and edge weights assigned as sequence similarity of the gene pairs. The homology graphs from all pairwise comparisons of species are combined into one, tripartite adjacency matrix. Using this initial gene graph, SAMap projects the three single-cell RNAseq datasets into a joint, lower-dimensional manifold representation. This joint manifold allows estimation of similarity between cells and genes across species. Note that a simplified version of such cell or gene similarity quantification is to only consider the one-to-one orthologs while ignoring the many-to-one, one-to-many, and many-to-many orthologous genes as well as the non-orthogonal relationship between non-orthologous but homologous genes, both of which are incorporated by the SAMap method. Next, the expression correlations between homologous genes are calculated in the initial joint manifold to reweight the edges of the gene-gene homology graph. Using the reweighted homology graph as the new input, SAMap then iterates until convergence to generate a final joint manifold. The expression correlation between homologous genes of the two species calculated across the joint manifold quantifies the similarity of the two genes' expression patterns. Homologous gene pairs with an expression correlation higher than 0.3 are deemed expression homologs - homologs that share similar expression patterns across mapped cells. Triads of mapped expression homologs from human, lemur, and mouse datasets were identified.

We then examined, for each expression homolog triads, whether the three gene pairs were assigned as orthologous genes by NCBI and/or Ensembl (see Table 7). We further examined for each expression homolog triad whether the lemur gene is named or unnamed in NCBI Annotation Release 101 of the mouse lemur genome assembly Mmur 3.0 (with only a locus identifier, e.g., "Loc__" or “orf”). Expression homolog triads were then categorized into three types (Fig. 3b). Type “named ortholog” refers to triads that consist of 3 orthologous gene pairs and the lemur gene is named accordingly. Type “unnamed ortholog” refers to triads of 3 orthologous gene pairs but the lemur gene is unnamed. Type “non-ortholog” refers to triads of less than 3 orthologous gene pairs, regardless of the naming status of the lemur gene. Table 4 lists the expression homologs detected in this study.

B cell receptor analysis

To improve the annotation of the B cell receptor (BCR) locus, we first used BLAST[155] with human BCR genes to find unannotated BCR genes in the mouse lemur genome and to extract variable and constant region genes, which we then used as a reference to extract and assemble the imunoglobulin (Ig) sequences from the lemur cell transcriptomic atlas. Cells identified as B cells and plasma cells in the SS2 datasets were selected for further analysis of their Ig sequence, which was assembled separately for each cell using BASIC[156] and a custom reference database of human sequences from IMGT (https://www.ebi.ac.uk/ipd/imgt/hla/) and mouse lemur Ig sequences (retrieved by BLAST searches with the human reference IMGT sequences of the mouse lemur Mmur 3.0 genome from NCBI). In addition, Ig sequences obtained by BLAST searches of the transcriptomes of a subset of mouse lemur atlas B cells were added to the custom reference database to further improve alignment. Constant regions from both the heavy and light chain contigs assembled by BASIC were aligned to these sequences using

BLAST and the best hit (with at least 80nt of overlap) was used to assign the isotype (i.e. *IGA*, *IGG*, *IGM*, or *IGE* for the heavy chain, and *IGK* or *IGL* for the light chain) for each cell. Putative V and J gene families and the *CDR3* sequences from both the heavy and light chain contigs were identified using IgBlast[157]. In some cases, BASIC was not able to generate a contig for the heavy and/or light chain, therefore the isotype was not assigned for these cells (52 and 10 cells for the heavy and light chains, respectively). In cases where BASIC was unable to assemble a single continuous contig from both constant and variable region ends of either the heavy and/or light chain, we submitted to BLAST/IgBLAST the two contigs constructed from each end (94 and 100 cells for heavy and light chains, respectively) or the only constructed contig from one end (147, 1, 28, and 2 cells with only heavy chain constant, heavy chain variable, light chain constant, and light chain variable contigs, respectively). In rare cases, the contigs constructed from the variable and constant region ends from a single cell returned different hits. Constant region isotypes were not assigned for such cells with different BLAST hits (16 and 13 cells for heavy and light chains, respectively). Similarly, V gene families were not assigned for cells with different IgBLAST hits (alignment quality V score >100) from the constant and variable contigs (3 and 59 cells for heavy and light chains, respectively). These discrepancies may be caused by doublets of B cells/plasma cells or reflect an actual gene rearrangement and dual expression. *CDRH3* lengths were calculated as the number of amino acids between the canonical C at the 5' end of the sequence and the 3' sequence WGXG where X is any amino acid. *CDRL3* (including lambda and kappa chains) lengths were calculated as the number of amino acids between the canonical C at the 5' end of the sequence and the 3' sequence FGXG or WGXG where X is any amino acid.

Because somatic mutation patterns in *IGV* genes can render a single V gene indistinguishable from separate but closely related alleles, we estimated a minimum number of V

genes based on the number of V loci that appeared in the current assembly of the genome. The number and orientation of these genes was determined manually by examining the results from blasting expressed V, J and C genes to the genome. Long read sequences covering this region would help determine the true number of V gene loci.

Clonal lineages were identified as groups ($N \geq 2$) of cells within a single individual with identical *CDRH3* and *CDRL3* lengths, the same light chain isotype and *CDR3* identities of at least 80% for both the light (*CDRL3*) and heavy (*CDRH3*) chains.

MHC gene expression analysis

Methods to examine mouse lemur MHC gene structure, extract MHC gene expression from the atlas, and re-annotate the MHC genes are detailed in an accompanying paper[50]. In brief, raw fastq files from 10x scRNA-seq of blood from each of the individuals were mapped against an MHC reference sequence derived from the BAC sequence[51], as well as the known expressed mouse lemur class I Mimu W01-04[158] using bowtie2. The mapping results were assessed for mismapping of reads, allelic variation and possible presence of additional genes, by manual inspection using the Integrative Genomics Viewer (IGV)[159] and Geneious Prime 2021.2.2. The fastq files were also ‘probed in silico’ by searching for reads containing sequences specific to the known genes. This confirmed the absence of expression of particular MHC genes. For the analysis of expression levels, a reference specific to each of the individuals was used to map the reads from the raw fastq files for each tissue using bowtie2. The sequence used was restricted to the final 600 bp comprising primarily the 3’UTR to avoid complications of potential recombinant sequences. Manual inspection of the results from the blood 10x scRNA-seq files was used to determine a mapping quality (MAPQ) threshold for each gene. The Sequence Alignment Map (SAM) file from the mapping was converted to a BAM file and then divided into

individual BAM files for each gene. These individual files were then filtered to remove reads below the MAPQ threshold. For the remaining reads, the cell barcode and UMI were counted. The expression level is normalized as read counts/10000 UMIs and then natural log transformed. Counts for each MHC gene (raw and normalized) are available in the metadata for every h5ad file in Figshare (https://figshare.com/projects/Tabula_Microcebus/112227).

Chemokine ligand and receptor expression analysis

A list of human chemokine receptors was compiled from the literature[160,161] and their cognate ligands were obtained from CellPhoneDB[162]. Of the 25 identified receptors, a corresponding lemur ortholog annotated in NCBI was found for all except *CCR2* (Table 6). Of the 48 cognate ligands, a corresponding lemur ortholog was identified for 34 (Table 6). The expression level of each of the lemur orthologs across all cell types in the 10x dataset is summarized in Fig. 5a and Fig. S6-7.

Cross-organ immune cell analysis

Immune subpopulations were identified and annotated by systematic subclustering of the lymphoid and myeloid compartment in each tissue for every individual, then adjusted by inspection using cellxgene after integration of tissues across all individuals to ensure consistency of cell type labeling, as described in the accompanying Tabula Microcebus manuscript 1[6]. Clusters branching off the main group were labeled with a differentially-expressed gene (e.g., neutrophil (*CCL13*+), neutrophil (*IL18BP*+), B cell (*SOX5*+)) and cells expressing proliferative markers (*MKI67, TOP2A*) were appended ‘PF’ (e.g., B cell (PF)). For macrophages, their identities as tissue-resident macrophages based on published marker genes (see Table 1 in the accompanying Tabula Microcebus manuscript 1[6]) was indicated by appending the corresponding

name (e.g. macrophage (Kupffer cell), macrophage (microglial cell)), given that clear distinction from monocyte-derived macrophages was challenging (with the exception of lung tissue-resident alveolar and monocyte-derived interstitial macrophages which were confidently distinguished based on canonical markers and labeled as such). Identification of differentially-expressed genes for each subpopulation was performed using the Wilcoxon rank-sum test, selecting genes with log fold change >=1 and *p*-value <0.05, after adjustment by the Benjamini-Hochberg correction.

For the cross-organ monocyte/macrophage analysis, all granulocyte-monocyte precursors, monocytes, and macrophages from the atlas were extracted for further analysis. Data were integrated across the four lemur individuals and then across the scRNA-seq methods (10x and SS2 datasets) using the FIRM algorithm[163] to correct for batch effects. In this integrated UMAP, monocyte populations co-clustered across tissues, whereas macrophage populations generally separated by tissue. We therefore tried additional FIRM integration across tissues, however tissue-specific separation of macrophage types and L2's bladder monocytes remained; hence we did not perform tissue-level FIRM integration for the final UMAP to avoid potential computational bias from overcorrection. We then examined expression of known monocyte and macrophage markers reported in the literature as well as the distribution of monocytes and macrophages from each tissue in the FIRM-integrated UMAP (Fig. S10d-f). In addition to the tissue-specific/resident populations highlighted in Fig. 6e, we found that pancreatic and heart macrophages also formed separate populations; however these were excluded from further analysis because they likely resulted from technical issues: the differentially-expressed genes for pancreatic macrophages were found to be broadly expressed in other cell types of the same tissue (signal spreading), and the heart sample had overall fewer transcripts per cell (lower quality).

The neutrophil developmental trajectory was based on embedding of neutrophils in the FIRM-integrated UMAP of the entire atlas (as described in the accompanying manuscript[6]). This resulted in co-clustering of neutrophils by individual and tissue, allowing for recapitulation of the developmental trajectory. We also tried FIRM integration of neutrophils alone (by individual and scRNA-seq method), however that resulted in separation of neutrophils by tissue and individual, which was largely driven by batch effect (no biologically meaningful differentially-expressed genes were identified across most clusters). This suggests that neutrophils are more molecularly homogeneous across tissues compared to other cell types such as macrophages. The trajectory was obtained by an in-house algorithm that detects the density ridge of the cell distribution on the FIRM integrated UMAP embedding, as described in the accompanying Tabula Microcebus manuscript 1[6], with the direction of the trajectory assigned manually based on expression of neutrophil maturation markers. Similarly to the neutrophils, the FIRM-integrated UMAP of B cells and plasma cells showed global separation of plasma cells and B cells, however further cell separation was driven by batch effect. In the atlas-wide UMAP, the clear separation between B cells and plasma cells precluded further trajectory analysis.

Endometrial cancer analysis

To compare the novel lung epithelial cell cluster in L2 (retrospectively identified as endometrial tumor cells metastasized from the uterus) with all other cell types of the atlas, we examined the correlation scores (Fig. 7i) and UMAP embedding (Fig. 7h) of their gene expression profiles, by the methods described in the accompanying Tabula Microcebus manuscript 1[6]. Here, we extracted results of the metastatic tumor cell type. In brief, to calculate the cell type pairwise correlation scores with the metastatic tumor cell type, atlas data was first integrated across individuals, tissues, and scRNA-seq methods (10x and SS2) using FIRM[163] and

FIRM-generated principal component coefficients were calculated for each cell. The coefficients were then averaged across all cells of a cell type and used to calculate the Pearson's correlation scores between every atlas cell type and the metastatic cells. The lung cell type in L2 that is a hybrid of metastatic and AT2 cells (which could be doublets of the two cell types) was excluded from the analysis. To generate the cell type UMAP (Fig. 7h), gene expression levels were averaged across cells for each cell type (10x dataset, excluding low quality cell types and ones represented by <4 individual cells, expression levels were normalized (0 to 1 scale) to the maximal value of each gene across all cell types, and the normalized cell type gene expression matrix projected onto a 2D space with cosine distances between pairs of cell types as input.

Differential gene expression analysis was performed on the lung metastatic cells versus all other lung epithelial cells and on uterine *FXYD4+ MUC16+* epithelial cells versus all other uterine epithelial cells (10x datasets) using the Wilcoxon rank-sum test (*p*-value <0.05, after adjustment by the Benjamini-Hochberg method), and representative examples selected for Fig. 7j.

Adipocyte analysis

Adipocytes and adipo-CAR cells were extracted from the FIRM-scaled and integrated data of the entire atlas (10x and SS2 datasets, see accompanying Tabula Microcebus manuscript 1[6]. The top 3,000 highly variable genes in the FIRM-transformed gene counts table of adipocytes and adipo-CAR cells were selected, and dimensionality reduction by PCA performed (top 13 PCs) to generate a 2D UMAP of adipocytes and adipo-CAR cells. Differential gene expression analysis on the *UCP1*hi versus *UCP1*lo adipocyte populations (L2 and L4, 10x dataset) was performed using Wilcoxon rank-sum test (*p*-value <0.05, after adjustment by the Benjamini-Hochberg method), and representative genes selected for Fig. 7n. Similarly, differential gene

expression analysis was performed between the adipocytes of each fat depot of L2 (BAT, MAT, SCAT, GAT; 10x dataset), and the top 10 genes enriched in each depot selected for Fig. S12e.

Most of the adipocytes in the atlas were isolated from fat depots, for which the tissue dissociation protocol was designed to enrich for the stromal vascular fraction (SVF) and exclude adipocytes (Supplementary Methods in accompanying Tabula Microcebus manuscript 1[6]). Most were from L2 (Fig. S12a),whose adipocytes in fat depots surrounding several tissues (e.g., kidney, spine, uterus) were generally small, predominantly multilocular, densely stained, and mitochondrial rich (Fig. S12d), features of brown or beige adipocytes in human and mouse, and intermingled with small, unilocular adipocytes with a single lipid droplet, which resemble white adipocytes. In contrast, L3 and L4 adipocytes were generally larger and most were unilocular (Fig. S12d). These may be harder to capture by current scRNA-seq protocols so may have contributed to the lower yield of adipocytes for L3 and L4.

Analysis of primate (human-lemur)-specific genes

The list of human and lemur orthologous genes with no corresponding mouse ortholog (Table 7) was compiled by merging human, mouse lemur, and mouse homology assignments from NCBI, Ensembl, and Mouse Genome Informatics (MGI) databases by a similar method used to compile the list of one-to-one-to-one gene orthologs for the comparison of cell types across the three species in the accompanying Tabula Microcebus manuscript 1[6]. We began by compiling all human protein-coding genes annotated in NCBI (human taxonomy ID: 9606), then merged the corresponding mouse lemur and mouse orthologs from NCBI (gene_info.gz and gene_orthologs.gz from https://ftp.ncbi.nlm.nih.gov/gene/DATA/, February 2020). We next added Ensembl gene ID numbers, gene names, and lemur/mouse ortholog assignments from Ensembl Biomart (Ensembl Genes version 99, February 2020), using the Ensembl gene ID

(variable 'Gene_stable_ID') for each NCBI gene ID (variable 'NCBI_gene_ID ') in Ensembl Biomart. MGI mouse gene ID numbers, gene names, and ortholog assignments (none provided for lemur) from Jackson Laboratory (HOM_MouseHumanSequence.rpt from http://www.informatics.jax.org/downloads/reports/, Feb 2020) were added using the MGI homology ID (variable 'HomoloGene_ID') attributed to each NCBI gene ID (variable 'EntrezGene_ID') in the MGI database. The Online Mendelian Inheritance of Man (OMIM) genetic disorder phenotype associated with each human gene (genemap2.txt from https://www.omim.org/downloads January 2022, variable 'Phenotypes') was added using the gene name (variable 'Approved_Gene_Symbol') in the OMIM database.

A human gene was identified as sharing an orthologous gene with lemur if at least one such assignment was made by either Ensembl or NCBI, and as sharing an orthologous gene with mouse if at least one such assignment was made by NCBI, Ensembl, or MGI. Conversely, a human gene was identified as sharing no orthologous gene with the other species if no such assignment was made in any of the two (lemur) or three (mouse) databases. The resultant 546 human and lemur-specific genes corresponded to 388 unique lemur Ensembl gene IDs and to 429 unique lemur NCBI genes IDs, 404 of which had detected transcripts in the atlas. Expression profiles of the 429 primate (human-lemur)-specific genes in all molecular cell types across the atlas (10x dataset, Fig. S8) were viewed in heat maps and dot plots and qualitatively categorized by inspection as enriched (higher or restricted expression), or selectively depleted, in one or more tissues/organs or compartments (Table 8). We also noted the cell types within the enriched tissues/organs and compartments that showed selective expression. Expressed genes that did not show any of these expression patterns were categorized as 'not enriched in any category'.

Gene set enrichment analysis of the 546 primate (human-lemur)-specific genes was performed using gprofiler2 in R[164]. Gprofiler2 finds over-representation of gene sets (relative to all human annotated genes) in Gene Ontology terms, biological pathways, regulatory DNA elements, human disease gene annotations, and protein-protein interaction networks. All default parameters were used to run the analysis (e.g. user_threshold = 0.05, correction_method = "g_SCS"). Enrichment results were tabulated to showcase the overlapping genes in each enriched category (Table 9).

FIGURE LEGENDS

Figure 1. Organism-wide scRNA-seq uncovers new mouse lemur genes

a. Schematic of the analysis used to identify all transcriptionally active regions (TARs) of the genome from 10x scRNA-seq data and to characterize the differentially-expressed, unannotated TARs (uTARs).

b. Pie chart of panel **a** showing fraction of the genome (percent of base pairs in Mmur 3.0 assembly) that is aTARs, uTARs, or transcriptionally inactive.

c. Stacked bar plot showing for each cell type in altas (10x dataset) the median percentage of transcript reads categorized as aTARs, differentially-expressed uTARs (DE uTAR, see Methods), and non-differentially-expressed uTARs (non-DE uTAR). Cell types are ordered by their designation number (see accompanying Tabula Microcebus manuscript 1[6]) and color coded by compartment. Examples of cell types with high proportions of DE uTARs are indicated by their designation number - cell type annotation (tissue).

d. UMAP visualization of mouse lemur male germ cells from testis (lemur L4, 10x dataset), with embedding based on expression of either the annotated genes (top, see also Sections 3 and 6 of accompanying Tabula Microcebus manuscript 1[6]) or uTARs alone (bottom). Each datapoint (cell) is color coded by its annotated stage of spermatogenesis (color code in panel g). Black lines with arrowheads indicate pseudo-time developmental trajectories, and fine gray lines show alignment of each data point to the trajectory.

e. Comparison of male germ cell pseudo-time trajectory coordinates based on expression of annotated genes (x-axis) or of uTARs (y-axis). Dashed black line, 1:1 relationship; *r*, Pearson's correlation coefficient.

f. Number of annotated genes (top) and uTARs (middle) expressed, and percent uTAR reads of total TARs (bottom), in each male germ cell, ordered along the pseudo-time trajectory as in panel d (bottom).

g. Dot plot showing average expression of selected DE uTARs (names based on subsequently identified sequence homology) across the indicated male germ cell types ordered by developmental stage. Dot fill indicates average expression level of the uTAR in the cell population, and dot size indicates percent of cells in the population with detected uTAR transcript. Identical gene names indicate multiple uTARs aligned to the same gene in another species.

h. Previously unannotated or incorrectly annotated immunoglobulin genes identified by scRNA-seq of B and plasma cells. Schematics of mouse lemur (top), human (middle) and mouse (bottom) immunoglobulin loci for heavy chain (*IGH*, left) and kappa (*IGK*, center) and lambda (*IGL*, right) light chains, located on the forward (fwd) or reverse (rev) strand of the indicated chromosomes (chr) and color-coded as in key. Top lemur line shows currently annotated gene order (NCBI Annotation Release 101, Mmur 3.0); line below shows revised annotation based on BLAST homology and transcript alignment of our atlas data (SS2 dataset). Filled boxes, constant (A, E, G, D, M for *IGH*, C for *IGK* and *IGL*), variable (V), joining (J) and diversity (D) regions; open boxes, pseudogenes. Numbers above V regions show number of functional V genes in region, numbers in parentheses estimate V pseudogenes (based on lack of transcripts aligning to gene); number of V genes for mouse lemur is approximate and varies between individuals. Gray dashed box, second V cluster ~5Mb downstream of constant region in *IGH* locus (main cluster is upstream of constant region) which may be an assembly error. Arrows indicate genes oriented opposite to direction of constant regions, and those with numbers indicate subset of those genes

that are flipped. Table below mouse lemur locus indicate number of expressed alleles for each constant region isotype in each lemur individual analyzed. Human and mouse data from previous studies[28,165]; humans have 7-11 JC clusters in *IGL* locus[28] (7 gene haplotype shown).

i. Barplot showing fraction of atlas B and plasma cells (SS2 dataset) of each immunoglobulin heavy chain (top) and light chain isotype (middle), and by their V_H family member usage (bottom), assigned from their scRNA-seq expression profiles and colored as in panel h and gray indicate unassigned isotype (see Methods); data are shown separately for each individual lemur (L2, L4). Fractions for heavy chain are also shown separately for organs with ≥5 cells. N, number of cells analyzed.

Figure 2. SICILIAN analysis of organism-wide scRNA-seq defines mouse lemur gene structures and splicing

a. Overview of mRNA splicing analysis based on SICILIAN. Bars, exonic sequences; lines, intronic sequences

b. Splice junction categories based on annotation status. Category A: Previously annotated splice junction. Category B: Novel junction between two annotated exon boundaries (e.g., novel exon skipping events). Category C: Novel junction between annotated exon boundary and unannotated location within gene. Category D: Novel junction between two unannotated locations within gene. Category E: Novel junction outside annotated gene boundaries.

c. Percent of splice junctions, scRNA-seq reads, and average number of reads per junction for each junction category.

d. Percent of splice junctions in each category that are annotated in both human and mouse genomes (HM), only in human (H), only in mouse (M), or neither (NE).

e. Cell-type-specific alternative splicing of *MYL6*. Top, structure of *MYL6* transcripts (isoforms 1,2) from exon 5 (E5) to exon 7 (E7) (NCBI annotation, detail in Fig. S2a; arrowhead, transcription direction 5'→ 3'); alternatively spliced exon 6 (E6) is red. Middle, mapped scRNA-seq read buildups for *MYL6* at E5 to E7 in indicated cell types; connecting arcs show number of sequence reads across the junction. Bottom, percent of isoform 1 (cyan) and isoform 2 (orange) in indicated cell types (combined 10x and SS2 scRNA-seq reads from all individuals); cell types across atlas are shown in Fig. S2a.

f. Barplots as in panel e showing compartment-selective (endothelial, stromal only) alternative splicing of *CAST*. Isoform 1 (cyan), exon 7 included; isoform 2 (orange), exon 7 excluded. Individual cell types in Fig. S2b.

g. Barplots showing developmentally-regulated alternative splicing of *MLF1, TPD52L2, FAM92A,* and *CHPT1* during spermatogenesis and in comparison with atlas-wide non-germ cells. NA, no spliced transcript of gene in corresponding cell type. Dashed line, insufficient reads covering corresponding region of transcript. Individual cell types in Fig. S2c-f.

Figure 3. SAMap analysis of atlas scRNA-seq data aids mouse lemur gene annotation

a. Percent of genes that are named (with gene symbol), unnamed (only a locus identifier, e.g., "Loc__" or "orf", and a suggested gene description), and uncharacterized (unnamed and with no gene description) in mouse lemur (left; genome assembly Mmur 3.0, NCBI annotation release 101), human (middle; assembly GRCh38.p13, NCBI annotation release 109), and mouse (right; assembly GRCm38.p6, NCBI annotation release 108), shown for all genes in genome (all) or for protein-coding genes (PCG) or non-protein-coding genes (nPCG).

b. Schematic of human-lemur-mouse expression homolog triad types based on similarity of their cell type specific expression profiles, genome-assigned orthology among the genes (according to

NCBI and Ensembl ortholog databases), and naming status of mouse lemur gene. Solid line, expression homologs with previously assigned gene orthology; dashed line, expression homologs without assigned orthology. Left (green): expression homolog triad involving assigned orthologs in all three species, and lemur ortholog is named accordingly. Middle (pink): expression homolog triad involving assigned orthologs in all three species, but lemur ortholog remained unnamed. Right (gray): expression homolog triad involving homologous (calculated by sequence homology[37]) genes that are currently not assigned orthologs for at least one of the species, regardless of the naming status of lemur gene.

c. Pie charts showing percent of named ortholog (green), unnamed ortholog (pink), or non-ortholog (gray) expression homolog triads identified in the lung and limb muscle scRNA-seq datasets, respectively.

d. Dot plot showing average expression ($\ln(UMI_g/UMI_{total} *1e4 +1)$, abbreviated ln(UP10K+1) and percent of cells (circle size) expressing the gene for five selected expression homolog triads of the "named ortholog" type across human, lemur, and mouse lung and limb muscle cell types. In each triad (boxed), human gene ortholog (blue) has "hu" prefix, lemur (black) has "le" prefix, and mouse (orange) has "ms" prefix. Cell types are ordered by compartment; Adv (adventitial), Alv (alveolar), MΦ (macrophage), cDC (conventional dendritic cell), pDC (plasmacytoid dendritic cell), FAP (fibroadipogenic progenitor cell), PF (proliferating). Diagram (right) shows example expression homolog triad (format of panel b; line color indicates expression homolog type).

e. Selected expression homolog triads of "unnamed ortholog" type. Note lemur genes have only locus identifiers and no gene symbol. In bottom two examples (*CD14*, *PALM*), two lemur orthologous genes with identical gene description were assigned by NCBI (i.e., monocyte

differentiation antigen CD14-like for *CD14*, paralemmin-1 for *PALM*), but only one is detected as an expression homolog (arrows), suggesting it is the true ortholog.

f. Selected expression homolog triad of "non-ortholog" type that likely represents a missed orthology assignment. Note lemur gene is assigned as an ortholog of human *ALDH1A1* but not of mouse *Aldh1a1* in NCBI or Ensembl.

g. Selected homolog triads of "non-ortholog" type. In each example, all relevant expression homolog triads are shown (including corresponding "named ortholog" triads); asterisk indicates the outlier non-orthologous gene. For example, three *RAMP* expression homolog triads are detected: two "non-orthologs" (i.e., hu *RAMP2* - le *RAMP1** - ms *Ramp2*, hu *RAMP3** - le *RAMP2* - ms *Ramp2*) and one "named ortholog" (hu *RAMP2* - le *RAMP2* - ms *Ramp2*). The shared expression pattern of lemur *RAMP1* and human *RAMP3* with *RAMP2* suggests that lemur RAMP1 and human RAMP3 have evolved to engage in similar physiological functions as the species-conserved RAMP2 or instead to modulate RAMP2-mediated ligand signaling in lung endothelial cells.

Figure 4. Structure and global expression pattern of mouse lemur MHC class I and II genes

a. Schematic of main human and mouse lemur MHC locus on chromosome (chr.) 6. Lemur genes depicted as annotated by NCBI (Annotation Release 101, Mmur 3.0). Filled rectangles, expressed genes; open rectangles, pseudogenes. Dark blue, classical MHC class I genes (high, ubiquitous expression); light blue, non-classical MHC class I genes (lower expression and/or tissue-specific expression); orange, MHC class II A genes; red, MHC class II B genes. Dashed box, haplotypic variability in human *DRB* gene number. Numbers below lemur *DRB*, *DQA*, and *DQB* indicate number of genes annotated by NCBI, though[50] suggests a single gene for each

family. Dashed lines, extended areas in genome assembly. g1, LOC105855356 in NCBI; g2, LOC105855357; g3, LOC105858107.

b. Schematic of MHC class I gene locus on mouse lemur chromosome 20. Top line, gene order as annotated by NCBI. The three genomic segments are separated by gaps (slanted lines) in Mmur 3.0 genome assembly. Second line, proposed reorganization of assembly based on rearrangement of three segments to match gene order of a sequenced BAC (third line; note gene content in Mmur 3.0 assembly and BAC differ due to haplotype variability in the individuals sequenced)[49]. Dashed lines, corresponding genes. Fourth line, proposed revision of structure and annotation of this locus based on above and the expression pattern of these genes in the mouse lemur cell atlas (panels c-d). Filled rectangles, expressed genes; open rectangles, pseudogenes. Dashed boxes, genes varying in either presence, copy number or expression status. *202* and *202P* are a pair of phylogenetically related genes annotated by NCBI, possibly representing a recent duplication; however, there was no evidence supporting them as separate genes in atlas expression data, suggesting *202P* is a pseudogene or assembly error. g4, LOC105855949 in NCBI; g5, LOC105855951; g6, LOC105870766; g7, LOC105870764; g8, LOC105870765; g9, LOC105870769; g10, LOC105870767; g11, LOC105870762.

c. Number of putative alleles for each MHC gene in the four lemurs (L1-L4), and number of genes annotated in the BAC and by NCBI (note some class I sequences differ only by a few base pairs among haplotypes (see accompanying MHC manuscript[50]), therefore the number of alleles in the panel represents our best estimate but may be inexact due to sequencing errors or other technical artifacts). Genes that exhibit copy number variation (C) or have at least one allele that is a pseudogene (P) are indicated below each panel. ?, insufficient number of reads for a reliable

allele count. NA, gene absent from BAC sequence. Based on sequence similarity, *168*, *W03* and *W04* could represent divergent allelic variants (or separate genes).

d. Dot plot of average expression of each MHC gene across all molecular cell types in the 10x dataset, combining all individuals. Cell types are grouped by compartment and numbered as in accompanying Tabula Microcebus manuscript 1[6].

Figure 5. Mapping expression of chemokine signaling and systemic immune responses

a. Dot plot of expression of selected chemokine receptors and their primary cognate ligands across immune cell types and other major interacting cell types in the atlas (10x dataset). (Figs. S6-S7 show epithelial cell types and all genes analyzed, and Table 6 provides full list of gene interactions). Boxes, cell types selectively expressing a receptor(s) (filled boxes) and cell types expressing its cognate ligand(s) (open boxes of same color). Many cell types show stereotypical expression of certain chemokine ligands and/or receptors across all lemurs and/or tissues analyzed; these may guide specific immune cell trafficking under normal physiology (see Section 4). We also noted inflammation and disease related expression of chemokines and receptors in certain individuals' cell types (cell type designations boxed in gray and illustrated for L2 in panel b).

b. Summary of cross-organ inflammatory processes identified for lemur L2, diagnosed with endometrial cancer (1) with metastatic spread to lung (2) and secondary uterine and pulmonary bacterial infection. L2 also had suppurative cystitis (3) and suspected inflammation of adipose tissues (4). These pathologies were accompanied by the local and systemic immune cell responses indicated, and the expressed genes that likely mediate the responses (cognate ligand and receptors are connected by a dash, *ligand — receptor*, and chemokines and receptors

featured in panel a are colored accordingly). Interactions of primary uterine tumor cells (asterisk) are based on transcriptomic profiles of L3, who also had an endometrial tumor (Section 6). For example, L2 showed an increase in circulating maturing neutrophils, as indicated by the left shift in L2's CBC test and detection of maturing neutrophils captured by scRNA-seq in peripheral organs (see section 5 and Fig. 6d). These immature/proliferating neutrophils as well as other immune progenitor cells (e.g., granulocyte-monocyte progenitors, GMP) are normally retained in the bone marrow by surrounding niche cells such as osteo-CAR cells through CXCL12-CXCR4 signaling[59–61] (5). Upon infection, maturing neutrophils are released from the bone marrow, presumably through down regulation of *CXCR4* and upregulation of *CSF3R, CXCR1/2* in these cells[62,63,167] (6). At the site of infection, two subtypes of activated neutrophils (designated as *IL18BP*+ and *CCL13*+ respectively) express high levels of chemoattractants to further recruit additional immune cells (7), such as *CXCL8* that recruits *CXCR1/2*-expressing cells (e.g., maturing neutrophils), *CCL5* (*RANTES*) that recruits *CCR5*-expressing cells (e.g., activated neutrophils and T/NK cells), downregulation of L-selected (*SELL/CD62L*) that facilitates extravasation of neutrophils from the blood into sites of inflammation, and in the bladder CSF3 that promotes proliferation of CSF3R-expressing maturing neutrophils and their recruitment through downregulation of *CXCR4*[77,83–86] (see Section 5, Fig. 6b-d, Fig. S9). In the bladder and adipose tissues, a specific population of inflamed/activated monocytes and macrophages also express high levels of cytokines *CSF3, CXCL8,* and *IL1A* (Section 5).

In adipose tissue, inflamed capillary cells and mesothelial cells expressed *CXCL9/10/11* (known to be induced by interferon-gamma which is sparsely expressed in L2's T cells) to attract *CXCR3*-expressing cells[75,76] (e.g., T/NK cells, B cells, and mature dendritic cells) (8), as well as non-myelinating Schwann cells expressing elevated levels of MHC class II genes (see Fig. 4d,

Fig. S4d), reported to occur in mouse under inflammatory conditions[168]. Finally, we identified molecular subtypes of lymphatic endothelial and reticular stromal cells in multiple tissues expressing *RARRES2, CCL21,* and/or *ACKR4* that jointly guide and attract immune cells into lymphatic tissue for antigen presentation and/or clearance (9) (Section 4). For example, the activated neutrophils found in multiple inflammation sites express markers of senescence including *CXCR4* and decreased levels of the "don't eat me signal" (*CD47*), facilitating their clearance from the circulation at the lymph node, bone marrow, and spleen[87,88].

Schematic created with BioRender.com.

Figure 6. Organism-wide mapping of immune cell development and activation

a. Integrated UMAP of atlas immune cells (10x and SS2 datasets from all tissues of all individuals, integrated by FIRM algorithm) with cells (dots) positioned by their overall transcriptional similarity and color coded by major immune cell type. Same UMAP highlighting proliferation state, sequencing method, individual, and tissue of origin is shown in Fig. S8a-d, and full integrated UMAP (including non-immune cells) is in Fig. 1g of accompanying Tabula Microcebus manuscript 1[6].

b. Neutrophil trajectory as in panel a with cells color coded by indicated neutrophil progenitor/subtype designation. Black line, pseudotime trajectory, with arrow indicating maturation direction, determined by expression of neutrophil markers; thin gray lines, individual cell alignments to trajectory.

c. Heatmap showing relative expression (normalized to stable maximum (99.5 percentile) of the gene along trajectory, 10x dataset) of orthologs of human neutrophil markers including indicated granulopoiesis genes, suggesting molecular trajectory of lemur neutrophil development and activation resembles that of human. Cells (columns) are ordered by pseudotime trajectory in b,

from progenitors (left) to (activated neutrophils (right) (see bar at top colored by cell type designation). Because non-activated neutrophils are highly abundant, only 10% are shown (uniformly subsampled along trajectory). *, marker genes without a mouse ortholog; +, marker genes known to not be expressed in mouse neutrophils. [], description of gene identified by NCBI as a gene locus ([CTSG], LOC105866609; [DEFA4L], LOC105881499; [DEFA1], LOC105881500; [CCL8], LOC105885739; [CCL4], LOC105881712).

d. Partitioning of immune cells along neutrophil trajectory (x-axis, cells colored by neutrophil cell type designation as in b) by tissue of origin (y-axis) and mouse lemur individual (L1-L4, merged in right panel and colored by individual). Light blue stripes, approximate trajectory range of non-activated, mature neutrophils, the main circulating (blood) neutrophil population in healthy individuals. Note L1 had neutrophil progenitors and immature cells in circulation (blood), indicating dysregulation of granulopoiesis, and L2 had maturing neutrophils in blood (classically known as a “left shift”) and activated neutrophils in peripheral and inflamed tissues, likely in response to infection. Gray dashed lines, organs not sequenced in that individual.

e. UMAP of atlas immune cells of monocyte and macrophage lineage including progenitors, integrated by FIRM across different tissues and individuals (10x and SS2 datasets). Note different populations of tissue-specific/resident macrophages and a population of activated monocytes (gonadal and bladder of L2). Inset (lower left), UMAP color coded by major cell types; see Fig. S10e-f for color coding by individual, tissue of origin, molecular cell type designation, and scRNA-seq method.

f. Dot plot showing expression of general macrophage markers and classical tissue-resident macrophage markers across the atlas macrophage populations. Note some markers are shared between multiple cell types (see Table 1 in the accompanying Tabula Microcebus manuscript

1[6]). The top differentially-expressed genes for each population are shown in Fig. S10b. [], description of gene identified by NCBI as a gene locus ([CD14L], LOC105862649; [CD163], LOC105869074; [CD209L], LOC105885453).

Figure 7. Cellular and molecular characterization of mouse lemur diseases and physiology

a. L2 intact lung showing metastatic endometrial tumor nodules on surface.

b, c. Hematoxylin-and-eosin-stained (H&E) sections of L2 lung showing tumor nodules extending into lung parenchyma (b)[74] and close-up of tumor cells (c).

d. H&E section of primary endometrial tumor in L2 uterus, which metastasized to lung.

e. H&E section of primary endometrial tumor in L3 uterus, which metastasized locally.

Scale bars: 1 mm (b), 50 μm (c), 50 μm (d), 20 μm (e). Full micrographs available online (Tabula Microcebus portal, https://tabula-microcebus.ds.czbiohub.org/).

f. UMAP of lung cells (L1-L4, 10x and SS2 datasets) integrated by FIRM and color coded by compartment (and color coded by cell type designation in Fig. S11a). Note unique cluster (arrow) of epithelial cells from L2 lung, identified as metastatic tumor cells from her endometrial tumor.

g. UMAP of L3 uterine cells (10x dataset) color coded by compartment (and color coded by cell type designation in Fig. S11b). Note cluster of uterine epithelial cells (arrowhead), designated as non-ciliated epithelial cell of uterus (*FXYD4+ MUC16+)*, presumed to be the primary endometrial tumor cells.

h. UMAP of all atlas cell types according to their overall transcriptional similarity (L1-L4, 10x dataset). Data points are cell types (unique combination of tissue and free annotation) and color coded by compartment, as described in accompanying Tabula Microcebus manuscript 1[6]. Inset

(boxed) shows close-up of indicated UMAP region with cell types color coded by compartment (circle fill) and tissue of origin (circle border), and labeled with their cell type designation. Note metastatic tumor cells (Met, arrow) from L2 lung and adjacent (molecularly similar) uterine epithelial cells (*FXYD4+ MUC16+*, arrowhead) from L3, identifying the likely uterine cell origin of the metastatic tumor. Cell type abbreviations are Lung: Met (uterine metastasis); Kidney: PCT (proximal convoluted tubule), PST (proximal straight tubule), LoH thin D (loop of Henle thin descending limb), LoH thin A (loop of Henle thin ascending limb), LoH thick A (loop of Henle thick ascending limb), DCT (distal convoluted tubule), CDp (collecting duct principal cell), CDi-α (alpha intercalated cell), CDi-β (beta intercalated cell), MD (macula densa).

i. Sina plot showing Pearson's correlation coefficients between L2 lung tumor cells and all other molecular cell types (color coded by compartment) in atlas (10x and SS2 datasets). Highest correlation coefficients include L3 uterine epithelial cells (*FXYD4+ MUC16+*).

j. Dot plot showing expression in lung and uterine epithelial cell types, separated by lemur individual, of endometrial (and ovarian) cancer (EC) serum marker genes and genes (indicated by *) known to be amplified, overexpressed, or mutated in EC with their cognate ligands (red), receptors (blue), and/or modulators (green)[119,122–124,169–175]. Also plotted are representative genes with enriched expression in lung metastasis compared to other lung epithelial cell types, and genes enriched in uterine FXYD4+ MUC16+ cells (presumed tumor) compared to other uterine epithelial cell types. [], description of gene identified by NCBI as a gene locus ([GZMBL], LOC105864431; [AOX2], LOC105856978; [AKR1B10L], LOC105857399 and LOC105860191).

k, l. FIRM-integrated UMAP of adipocytes and adipo-CAR cells (10x and SS2 datasets, see Methods), and colored by cell type (k) or expression of the indicated genes (l). Adipocytes

formed two main populations, distinguished by expression of classical white (e.g., *NNAT*, panel l left) and brown (e.g., *UCP1*, panel l right) adipocyte markers, and designated as *UCP1*lo and *UCP1*hi, respectively. Note the within-population heterogeneity of the *UCP1*lo population, which forms two subclusters in the UMAP that differ only in total gene and UMI levels per cell, and not in expression of any biologically significant genes (Fig. S12a-c).

m. Distribution of *UCP1*hi (pink) vs. *UCP1*lo (blue) adipocytes in indicated fat depots and organs with ≥5 adipocytes (10x and SS2 datasets, all from individual L2, except combined fat depots from L4). N cell, number of adipocytes detected in each depot/organ.

n. Dot plot of expression of indicated cell type markers and differentially-expressed genes in *UCP1*hi and *UCP*1lo adipocytes, adipo-CAR cells, osteoblasts, osteo-CAR cells, and chondrocytes (L1-L4, 10x dataset). Note classical brown adipocyte marker *CIDEA* and white adipokine *RBP4* (both indicated by *) are surprisingly equally expressed across all adipocytes. [], description of gene identified by NCBI as a gene locus ([ATP1A2], LOC105862687; [COX7A1], LOC105876884; [Uncharacterized 1], LOC105854963).

Figure 8. Expression patterns of primate-specific genes in mouse lemur

a. Proportion of the ~20,000 human protein-coding genes with identified ortholog(s) (by NCBI, Ensembl, and/or MGI) in lemur and mouse. The 546 (3%) human genes that share an ortholog only with lemur (human-lemur or “primate-specific” genes) correspond to 487 lemur genes (327 annotated by both NCBI and Ensembl, 102 by NCBI only, and 58 by Ensembl only). Expression patterns of the 429 NCBI-annotated lemur orthologs were examined and categorized.

b. Pie chart showing fraction of primate-specific genes (168) enriched (or depleted) in a specific tissue compartment. Genes enriched (or depleted) in >1 compartment are indicated as ‘cross-

compartment'. Note one-third (31%) of the genes are enriched (or specifically depleted) in the germline compartment.

c, d. Sina plots showing expression of representative primate-specific compartment-enriched (or specifically depleted for *FOXO3*) genes (c) and organ-enriched genes (d) across all molecular cell types in the atlas (10x dataset), with cell types (data points in figure) grouped by compartment (c) or by organ of origin (d). [], description of gene identified by NCBI as a gene locus ([TRGC10], LOC105878255; [AMY2BL], LOC105863954; [MT1EL], LOC105866478; [CARD18L], LOC105862464).

e. Dot plot of expression of primate-specific germ compartment enriched ("Germ enriched") and depleted ("Germ depleted") genes across the non-germ compartments (averaged across all cells in the compartment) and germ cell types (10x dataset). [], description of gene identified by NCBI as a gene locus ([H2BC12], LOC105858749; [Uncharacterized 1], LOC109730269; [AK1], LOC105869668; [HSFX4], LOC109730266; [SPANXN4], LOC105864720; [MT2A], LOC105866476, [MT2AL], LOC105866477, [MT1XL], LOC105866553, [RPL36AL], LOC105873222).

TABLES

Table 1. Differentially-expressed-uTARs and their homologous genes and cellular expression levels

Table 1a (DE uTAR characteristics) lists the 4003 unique differentially-expressed (DE) uTARs detected in the atlas. Columns include uTAR identifier (unique uTAR ID that appends genomic position, strandedness (+/- strand), and total count of aligned reads followed by a "-0", indicating the TAR region is a uTAR), fasta (uTAR location in the genome in the FASTA format, i.e.,

chromosome ID followed by the start and end locations), fastaPeak (full width at half maximum region around the absolute peak in coverage after Gaussian smoothing within the uTAR location), tissue_cell type_all (list of cell types that the uTAR is differentially-expressed in each atlas tissue), num_tissue_cell type (number of cell types the uTAR is differentially expressed), tissue_cell type_top (the cell type with the most significant p-value in the differential expression test), avg_logFC (log(e) fold increase of the expression in the indicated top cell type compared to other 10x cell types in the same tissue), p_val_adj (*p*-value of enriched expression in the indicated top cell type, adjusted by Bonferroni correction), pct.1 (percent of cells in the indicated top cell type expressing the uTAR), pct.2 (proportion of other 10x cells in the same tissue expressing the uTAR), blastShort (most frequent significant BLASTn result for the entire length of the uTAR), blastShort_peak (most frequent significant BLASTn result for the peak region in the uTAR). Table also indicates whether DE-uTAR nucleotide sequence has homology with a human gene that has no known mouse or lemur orthologs (is_H=TRUE), and whether DE-uTAR nucleotide sequence has homology with a human gene that has a mouse ortholog and no known lemur orthologs (is_HM=TRUE). The table also highlights DE-uTARs that overlap in genomic position with an Ensembl annotation that is missing in NCBI (is_EnsemblOnly=TRUE).

Table 1b (uTAR cellular expression) lists the identity of each cell (cell ID, tissue source, individual source, cell ontology, and free annotation, as described in https://tabula-microcebus.ds.czbiohub.org/whereisthedata) and its TAR characteristics including for each cell the percentages of reads (UMIs) detected by the TAR analysis that are annotated TARs (percent_aTAR) as well as unannotated TARs (percent_uTAR) separated by differentially-expressed uTARs (percent_uTAR_DE) and not differentially-expressed uTARs (percent_uTAR_nonDE).

Table 1c (cell x uTAR count table) lists the UMI count matrix of the DE-uTARs across all 10x cells, with each entry being a non-zero element of the matrix followed by the corresponding cell index (index_cell), uTAR index (index_uTAR), and number of UMI (n_reads). The uTAR and cell indexes are as indicated in Table 1a and 1b, respectively.

Table 2. Spice junctions and categories identified by the SICILIAN analysis

Table lists all splice junctions detected by SICILIAN. Each row corresponds to a distinct junction and columns are: junction_ID (unique name for each junction providing chromosome name, gene name, coordinate, and strand orientation for the 5' and 3' sides of the junction), num_reads_total (total number of reads mapped to the junction across all 10x and SS2 datasets), chromosome_refseq_ID (RefSeq ID for the chromosome), chromosome_name (chromosome name), junction_5SS (5' splice site coordinate), junction_3SS (3' splice site coordinate), gene_5SS (gene name for the 5' side of the junctions), gene_3SS (gene name annotation for the 3' side of the junctions), strand (strand orientation), conservation (conservation status with human and/or mouse as defined in Fig. 2d), category (annotation category as defined in Fig. 2b).

Table 3. Genes differentially spliced across cell types

Table of genes with statistically significant splicing differences across cell types in the same tissue for at least two lemurs, detected by MANOVA test using the 10x atlas dataset. Each row lists a unique combination of gene, tissue source, and lemur individual source. Additional columns include pval_zscore_adj (the p-value resulting from the MANOVA test) and num_cell_types (number of cell types analyzed with the MANOVA test in the given tissue and individual for this gene).

Table 4. Expression homolog triads of human, lemur, and mouse detected by SAMap

Table lists expression homolog triads detected by SAMap analysis separately on the lung and limb muscle cross-species datasets (listed in column "Dataset"). Each row represent an expression triad with its type ('Type_expression_homolog'), gene symbols in the three species ('huGene', 'leGene', 'msGene') including its species prefix, status of lemur gene annotation in NCBI ('is_leGene_Named'), number of gene pairs in the triads that are orthologs ('N_ortholog_pair') or non-orthologs ('N_non-ortholog_pair'), pairwise orthology relationship between species ('huleOrthology', 'humsOrthology', 'lemsOrthology'), expression pattern correlation score between gene pairs ('huleCorr', 'humsCorr', 'lemsCorr'), and the cell types with selective expression of the gene in respective species ('huCellType', 'leCellType', 'msCellType').

Table 5. Expression of immunoglobulin isotypes in atlas B cells and plasma cells

Table lists atlas B cells and plasma cells (SS2 dataset) with the immunoglobulin isotype identified by the scRNA-seq profile of each cell, including the expressed constant region isotype (heavy_constant_isotype, light_constant_isotype) and variable region family member (heavy_variable_family, light_variable_family) of heavy and light chains, as well as the complementarity-determining region (CDR3) amino acid sequence and length of the heavy (CDRH3_aa_sequence, CDRH3_aa_length) and light chains (CDRL3_aa_sequence, CDRL3_aa_length). Blank entries in these columns indicate cells for which the isotype could not be assigned (see Methods). Pairs or triplets of cells identified as part of the same clonal lineage, based on their CDRH3 and CDRL3 sequence similarity, are given the same number in column 'clone_number'.

Table 6. Chemokine receptor and cognate ligand genes

Table lists the analyzed lemur orthologs designated by NCBI ('receptor_annotated', 'ligand_annotated') of the 25 human chemokine receptors and their cognate ligands, retrieved from CellPhoneDB. Source of each curated ligand-receptor interaction is given under 'annotation_strategy' (CellphoneDB output, 'curated' indicates it was annotated by CellPhoneDB developers). For receptors with multiple ligands, the main ligand (when known) is noted under 'comments'.

Table 7. Human genes and their lemur and mouse orthology assignments

Table 7a (Summary of human orthology assignments and primate-specific genes) summarizes the orthology status of all human protein coding genes annotated in NCBI (19,966), beginning with the NCBI gene ID, synonym, gene description, and gene type. The presence of corresponding mouse lemur orthologs assigned by NCBI and/or Ensembl (as detailed in Methods) is indicated in columns 'HumanLemur_NCBI_HomologyType' and 'HumanLemur_Ensembl_HomologyType', respectively. Similar columns are provided for corresponding mouse orthologs assigned by NCBI, Ensembl, and/or MGI. For every human gene, column 'any_Lemur_homology' summarizes if there is at least one lemur ortholog assigned by NCBI and/or Ensembl and column 'any_Mouse_homology' summarizes if there is at least one mouse ortholog assigned by NCBI, Ensembl, and/or MGI. Human genes that are primate-specific (546) are identified in the column 'HumanLemurOrthogs_NotMouse' (1: 'any_Lemur_homology' = yes and ''any_Mouse_homology' = no, 0 for all other entries).

Table 7b (All human orthology assignments) lists all human protein coding genes annotated in NCBI (19,966) with corresponding NCBI gene ID, synonym, external database cross references (dbXrefs), chromosome location, map location, gene description, gene type, as well as

corresponding Ensembl gene ID, gene name and OMIM associated phenotype description. The presence of corresponding mouse lemur orthologs assigned by NCBI and/or Ensembl (as detailed in Methods) is indicated in columns 'HumanLemur_NCBI_HomologyType' and 'HumanLemur_Ensembl_HomologyType', respectively, with the corresponding NCBI gene ID, gene name, and synonym as well as Ensembl gene ID and gene name listed for lemur genes with an assigned ortholog in either of these databases. Similar columns are provided for the corresponding mouse orthologs assigned by NCBI, Ensembl, and/or MGI. Rows are duplicated for each human NCBI gene with more than one corresponding Ensembl gene, or that is assigned to more than one human/mouse ortholog.

Table 8. Primate-specific genes and their organism-wide expression patterns in the mouse lemur

List of mouse lemur genes (429) annotated in NCBI that are primate (human-lemur)-specific, starting with the corresponding NCBI gene ID, synonym, external database cross references (dbXrefs), chromosome location, gene type, and gene description. For each gene, the compartments and tissues that have enriched expression of the gene are listed in columns 'Enriched_compartment' and 'Enriched_tissue', and summarized in column 'Expression_pattern_summary'. Cell types showing selective expression in the enriched compartments/tissues are noted in column 'Enriched_celltype'. Compartments and tissues that have selectively depleted expression are indicated by brackets. The number of cell types (10x dataset, excluding cells labeled as mix, doublets or low quality) with mean expression of the gene greater than zero (column 'Ncelltypes_wMeanExpGreaterThan0'), as well as the number of cell types with mean expression of the gene greater than zero and expressed in at least 1% of the cells in that cell type ('Ncelltypes_wMeanExpGreaterThan0_ExpIn1PCTOfCells') are provided.

The order of genes appearing in Fig. S8a-b heat maps are provided in column 'Gene_order'. Primate-specific genes (aTARs) found using the TAR analysis pipeline (see Methods - uTAR analysis) are indicated in the 'Gene_inTARanalysis' column. For some genes (see 'Comments' column), the current NCBI synonym or gene ID is used in the lemur scRNA-seq dataset instead of the NCBI gene name.

Table 9. Gene sets enriched in primate-specific genes

Gene set enrichment analysis (gprofiler2[164] in R) for the 546 primate (human-lemur)-specific genes. Columns detailed in https://cran.r-project.org/web/packages/gprofiler2/vignettes/gprofiler2.html.

SUPPLEMENTARY MATERIALS

Figure S1. uTAR analysis enhances mouse lemur genome annotations and uncovers potentially novel transcripts.

a. Box plot of average silhouette coefficient values of the atlas datasets (separated by tissue and lemur individual, and sequencing channel) based on expression of annotated genes, aTARs, or uTARs. For each dataset, cells were grouped by their cell type designation and their silhouette value was calculated using the cell-cell distances in the 2D UMAP embeddings of either the gene, aTAR, or uTAR-expression space, and then averaged across allcells of the dataset (see Methods). Box, mean±s.d.; red triangles, L4 colon dataset (10x) shown in panel b.

b. UMAP visualization of mouse lemur colon cells (L4, 10x dataset) embedded according to the expression of the annotated genes (top) or uTARs (bottom). Datapoints (cells) are color coded by cell type designation (as in c).

c. Dot plot showing expression of selected DE uTAR (labeled by identified sequence homology) across L4 colon cell types. Identical gene names indicate multiple uTARs aligned to the same gene in another species.

d. Expression of selected sperm cell type markers in germ cells from testis (L4, 10x dataset) ordered by the pseudotime developmental trajectory calculated by uTAR expression (as in Fig. 1d-g.)

e. Percentage of genes detected by the TAR analysis (for indicated gene categories) as a function of the filtering threshold used to define cell type selective expression (i.e. TAR expression in any cell type ≥ e^threshold times that of the average of other cell types). Gene categories examined include: the top 100 (black), 2000 (dark gray), and 5000 (light gray) variably-expressed genes annotated in Mmur 3.0 NCBI annotation, all genes (blue), primate (human-lemur)-specific genes (yellow), and genes annotated in Mmur 3.0 Ensembl annotations but missing from NCBI (green).

f. Venn diagram of 4003 DE-uTARs with sequence homology to coding regions (>1 hit by DIAMOND blastp analysis) and/or non-coding regions (>1 hit by Infernal cmscan analysis), according to Nf-predictorthologs analysis.

Figure S2. SICILIAN analysis reveals cell type specific alternative splicing

Barplots showing percent of indicated splice isoforms of *MYL6* (a), *CAST* (b), *MLF1* (c), *TPD52L2* (d), *FAM92A* (e), and *CHPT1* (f) as in Fig. 2e-g but across all atlas cell types (designation color coded by compartment: epithelial, endothelial, stromal, lymphoid, myeloid, neural, and germ) that have at least 10 cells with spliced transcripts of the gene and a total of ≥300 reads of spliced transcripts of the gene. Exceptions were the spermatogonia of *TPD52L2* (1 cell, 8 reads), *FAM92A* (7 cells, 21 reads), and *CHPT1* (34 cells, 112 reads) and diplotene spermatocyte of *FAM92A* (37 cells, 50 reads), and *CHPT1* (88 cells, 148 reads). In the transcript

annotations at top, splice isoforms are labeled with the NCBI Refseq ID and colored to correspond with barplot, with exons affected by alternative splicing shown in red. Note for *CHPT1* (f), the two transcripts in dashed box cannot be distinguished using the atlas data due to insufficient reads covering dashed region of the transcript so reads of the two isoforms are combined in the analysis. Note *MYL6* and *CAST* show compartment-dependent patterns of splicing, and *MLF1*, *TPD52L2*, *FAM92A*, and *CHPT1* are differentially spliced in male germ cells (testes #250-256).

Figure S3. Additional examples of expression homolog triads of the "named ortholog" type

Dot plots as in Fig. 3d showing expression of selected expression homolog triads of the "named ortholog" type, identified separately from the lung (a) or limb muscle (b) datasets (see Methods).

Figure S4. Mouse lemur MHC locus and expression patterns

a. Comparison as in Fig. 4a of the main MHC class I and class II regions of mouse lemur with other primates and the mouse[45,49,51,176–178]. chr., chromosome on which the MHC regions are located indicated in parentheses. Genes colored and shaded as in Fig. 4a. Gray dashed box, genes with haplotypic variability in gene number or expression status. Cyan dashed boxes enclose mouse MHC haplotype T, M, and Q regions with expanded gene families.

b. Schematic representation of the three *Mimu-DRB* genes in current mouse lemur genome assembly (Mmur 3.0). *DRB1-10* is predicted to encode a full length DRB1 polypeptide, whereas *DRB1-1* and *DRB1-4* are incomplete and contain non-MHC sequences (gray shading). Note that the MHC sequences in *DRB1-1* and *DRB1-4* encode a functional DRB polypeptide, suggesting that these sequences were misassembled and belong together to form a second complete *DRB1* allele[50]. Thin vertical lines, positions that differ between the three sequences.

c. Sina plot showing summed expression of classical class I, non-classical class I and class II MHC genes, respectively and averaged across individual cell types. Specific cell types discussed in the text are colored.

d. Dot plot showing summed expression of classical class I, non-classical class I, and class II MHC genes, respectively averaged by cell type and individual. 'x', cell type not sequenced in the individual.

e. Sina plot showing ratio of summed non-classical to summed classical MHC class I expression. Cell types with ratio >1 are labeled.

Figure S5. Immunoglobulin expression and clonal analysis

a. *CDRH3* lengths (number of amino acids, aa) for lemurs L2 and L4 (SS2 data), as well as healthy humans and laboratory mice. N, number of cells for mouse lemurs and number of unique clones for humans and mice, including all isotypes. Human and mouse data courtesy of Scott Boyd's group and Tho Pham[58,179]. *CDRH3* lengths over 30 were rare in human and mouse and did not occur in the mouse lemurs analyzed and are therefore not displayed. Note *CDRH3* length in lemur is generally shorter than in human and more comparable to that of mouse. Though it is known that the *CDRH3* region length also varies with age, disease state, and B cell maturity to affect antigen-binding affinity, the functional relevance of inter-species variation remains unsettled[180].

b. B and plasma cell clones identified by their *CDRH3* sequence. Each clone is represented as a circle, with the outline color indicating the identity of the lemur in which the clone was found and the fill color indicating the heavy chain isotype(s) of the clone (colors as in Fig. 1h-i). All clones consisted of two cells except the larger circle (in spleen) which is a three-cell clone. Circles on an organ indicate organ in which the constituent cells of the clone were found, and

circles between organs indicate the two organs in which the constituent cells were found. Organ images were created by Mikael Häggström and Madhero88 and are in the public domain.

Figure S6. Expression of selected chemokines and receptors in related epithelial cells

Extension of Fig. 5a with dot plot showing average expression across all epithelial cell types in the atlas (10x dataset) of indicated chemokine receptors and their main cognate ligands. Note that in epithelium, *ACKR4* is selectively expressed on skin keratinocytes and is known in other organisms to scavenge soluble CCL19/21, creating a gradient that aids in migration of dendritic cells and Langerhans cells to the draining skin lymph nodes[67,69]. *CCL20*, in addition to its known expression in small intestine and colon enterocytes of human[181], is also expressed in epithelial cells of the skin, tongue and bladder.

Figure S7. Expression of all chemokines and receptors across all atlas cell types

Dot plot showing expression across all molecular cell types in atlas (10x dataset) of all chemokine receptors and cognate ligands annotated in NCBI Refseq Annotation Release 101 for Mmur 3.0 (Table 6). Cell types are ordered by compartment, then cell type designation number, then tissue. [], description of gene identified by NCBI as a gene locus ([CCL4], LOC105881712; [CCL8], LOC105885739; [CCL11], LOC105885804; [CCL13], LOC105859268; [CXCL4], LOC105860501).

Figure S8. FIRM-integrated UMAP of all atlas immune cells and comparison of different types of B lymphocytes

a-d. UMAP of all atlas immune cells as in Fig. 6a but color coded by proliferation state (a), scRNA-seq method (b), individual (c), and tissue of origin (d).

e. Dot plot showing expression in the indicated B lymphocyte lineage cells in atlas (top) of marker genes for B cells, plasma cells, and top differentially-expressed genes in the *SOX5*+ B cell population (identified in pancreas) compared to other B cells.

Figure S9. Further characterization of activated neutrophil subtypes

a. UMAPs of lung neutrophils (10x dataset) of the indicated individuals, with cells color coded by their molecular cell type designation. Note the two subtypes of activated neutrophils (*CCL13*+ and *IL18BP*+) of Fig. 6b-d cluster separately from the main population of lung neutrophils (non-activated) in L1 and L2; neither activated subtype is detected in L3 or L4.

b. Heatmap showing relative expression of the indicated differentially-expressed genes in each of the activated neutrophil subtypes compared to other activated and mature neutrophils. Bars at top show, for each neutrophil, its molecular cell type designation (bottom bar), ty individual lemur source (middle bar), and tissue source (top set of bars). Single cell expression levels are normalized to the stable maximum (99.5 percentile) of that gene across cells along the neutrophil trajectory. The mature (non-activated) neutrophils are highly abundant so ones shown are 1/5 subsamples of non-activated neutrophils at late stage (>0.7) of the pseudotime trajectory (corresponding approximately to cells within blue stripes in Fig. 6d). [], description of gene identified by NCBI as a gene locus ([IFITM3L], LOC105874071; [Uncharacterized 1], LOC105867541; [CCL2L], LOC105859340 and LOC105885684; [CCL13], LOC105859268; [Uncharacterized 2], LOC105856756; [CCL3L], LOC105881608).

Figure S10. Further characterization of the monocyte and macrophage lineage

a. Expansion of dot plot in Fig. 6f showing expression of classical marker genes for hematopoietic precursors, granulocyte-monocyte progenitors, monocytes, macrophage, and tissue-resident macrophage markers across the atlas immune cells in the monocyte/macrophage

trajectory. Note some markers are shared between multiple cell types (see Table 1 in the accompanying Tabula Microcebus manuscript 1[6]). [], description of gene identified by NCBI as a gene locus ([CD14L], LOC105862649; [CD163], LOC105869074; [CD209L], LOC105885453).

b. Dot plot showing expression of differentially-expressed genes of the indicated tissue-resident macrophage populations in the atlas, compared to all other macrophage populations. [], description of gene identified by NCBI as a gene locus ([LOC105878881], CD300C; [LOC105876782], HLA-DRB1L; [LOC105869752], HLA-DQA2; [LOC105869753], CCL3L; [LOC105872655], TLL; [LOC105866341], SIGLEC8; [LOC105882132], SIGLEC7; [LOC105873562], FCGR3AL; [LOC105871481], KRT76L; [LOC105863040], PRXL2A).

c. Dot plot showing expression of the top differentially-expressed genes of the separately-clustered monocytes from L2 bladder and gonadal fat (see Fig. 6e), compared to the other atlas monocytes. [], description of gene identified by NCBI as a gene locus ([Uncharacterized 1], LOC105869025).

d. FIRM-integrated UMAP as in Fig. 6e showing expression in monocyte/macrophage trajectory of classical and non-classical monocyte markers as well as M1 and M2 macrophage markers for human and mouse. Note lemur monocytes and macrophages show no clear pattern of these markers. For reference, human classical monocytes (*CD14++*, *CD16-*) express higher levels of *CD14, CCR1, CCR2, CXCR1, CD11B/ITGAM, SELL/CD62L,* and *CD163* whereas non-classical monocytes (*CD14+*, *CD16++*) express higher levels of *CD16/FCGR3A, CD11C/ITGAX, CD64/FCGR1A, CX3CR1*, with intermediate monocytes expressing intermediate levels of those markers[108,182,183]. Mouse pro-inflammatory monocytes (*Ly6c*hi) – similar to human classical monocytes – express higher levels of *Ly6C, Ly6G/Gr-1, Ccr2, Cd43/Spn*, whereas patrolling

monocytes (*Ly6c*lo) – similar to human non-classical monocytes – express higher levels of *Cx3cr1* and *Sell/Cd62l*.

In both species, M1 macrophages express higher levels of *IL12A/B, IL23A, IL1B, IL6, and TNF*, whereas M2 macrophages express higher levels of *IL10 and TGFB1*[107,108]. Genes not annotated in the lemur genome include: *CCR2, CD64/FCGR1A, CD11C/ITGAX, Ly6c1, Ly6c2, Ly6G/Gr-1, CD43/SPN*. [], description of gene identified by NCBI as a gene locus ([CD14L1], LOC105862649; [CD14L2], LOC105862489; [CD16L], LOC105873562; [CD163], LOC105869074).

e. UMAP of atlas immune cells in monocyte/macrophage trajectory, color coded by individual source (left), scRNA-seq method (center), and tissue source (right).

f. UMAP of atlas immune cells from monocyte/macrophage trajectory, separately highlighting cells from each tissue, with cells color coded by tissue (left), individual source (center), and molecular cell type designation (right).

Figure S11. Further characterization of endometrial tumor metastasis cells in lung and uterine *FXYD4+ MUC16+* non-ciliated epithelial cells

a,b. UMAPs of lung (a) and uterine (b) cells from atlas, as in Fig. 7f-g, with cells color coded by epithelial cell type designation. Cells in gray are non-epithelial cells.

c-f. Same UMAPs of lung (c, e) and uterus (d, f) with cells colored to show expression of *OXTR* (c, d) and *MUC16* (e, f).

Figure S12. Mouse lemur adipocytes and expression patterns of adipocyte differentially-expressed genes, selected adipokines, and their receptors

a. FIRM-integrated UMAP of adipocytes and adipo-CAR cells (10x and SS2 datasets, see Methods) as in Fig. 7k-l with cells color coded by cell type designation, scRNA-seq method, individual lemur source, and tissue source, respectively.

b. UMAP as above color coded by expression level of indicated adipocyte markers (*ADIPOQ, CIDEC*) and differentially-expressed genes in *UCP1*lo (*CHIT1, APOE*) and *UCP1*hi (*FABP3, KCNK3*) adipocytes.

c. UMAP as above color coded by the number of scRNA-seq reads per cell (UMIs, 10x dataset; transcripts, SS2 dataset, left) and number of genes detected per cell (right). Note the heterogeneity of *UCP1*lo population, which forms two subclusters distinguished by total read per cell and genes detected per cell, but not by any biologically significant differentially-expressed genes.

d. Micrographs of H&E-stained sections of fat tissues from L2 (left) and L4 (right) that are near the kidney (top) and paraspinal muscle (bottom). Scale bar, 50 μm (all panels). Full micrographs available online (Tabula Microcebus portal, https://tabula-microcebus.ds.czbiohub.org/).

e. Dot plot of expression of the top 10 differentially-expressed genes in each of the four fat depots: BAT, interscapular brown adipose tissue; GAT, perigonadal; MAT, mesenchymal; SCAT, subcutaneous (L2, 10x dataset). [], description of gene identified by NCBI as a gene locus ([Uncharacterized 1], LOC105856764; [Uncharacterized 2], LOC105867540; [COX7A1], LOC105876884; [Uncharacterized 3], LOC105867541; [MT2A], LOC105866476; [MT1E], LOC105866554; [CTRB1L], LOC105875474; [MAGEB16L], LOC105877758; [PRSS1L], LOC105873340; [IGLL1], LOC109729893; [IGLL5], LOC105882024; [RPS3],

LOC105862350; [RPS20], LOC105874908; [RPS27L], LOC109731171; [RPL32], LOC105861123; [RPLP1], LOC105859117; [RPS15A], LOC105857549; [RPL29], LOC105863618; [FTL], LOC105870251).

f. Dot plot of expression of adipokines *LEP* and *ADIPOQ* as well as their receptors across the atlas cell types (L1-L4, 10x dataset). Note abundant and specific expression of *ADIPOQ* but lack of *LEP* expression in adipocytes. Curiously, the *AKR1B1+* kidney loop of Henle cells, mesothelial cells, and some vascular-associated smooth muscle cells show *LEP* expression, although at very low levels. In contrast, *LEPR* shows expected expression in various cell types including tendon cells, fibroblasts, and endothelial cells, and high *LEPR* expression is found in mesenchymal progenitor cell types such as osteo-CAR and adipo-CAR cells. Also note ubiquitous expression of *ADIPOR1* across almost all atlas cell types, and enriched expression of *ADIPOR2* in sperm lineage cells and adipocytes[40].

Figure S13. Atlas-wide expression of primate (human-lemur)-specific genes

a-b. Heat maps of gene expression of the 429 primate (human-lemur)-specific genes across all molecular cell types in the atlas (10x dataset). Cell types are ordered by compartment, then cell type designation number, then tissue source (a), and by tissue source, then compartment, then cell type designation number (b). Genes are ordered (left to right) by compartment-enriched (or depleted) genes, broadly-expressed genes, and genes not detected in the atlas.

c-h. Dot plots showing expression of compartment-enriched (or depleted) primate (human-lemur)-specific genes across all molecular cell types in the atlas (10x dataset). Genes are organized by enriched compartment (e.g. epithelial (c), endothelial (d), stromal (e), immune (f), neural (g), germ (h)), and genes are shown in more than one plot when enriched in multiple compartments. See Table 8 for full list of the primate-specific genes and their description.

REFERENCES

1. Hu, C.-K. & Brunet, A. The African turquoise killifish: A research organism to study vertebrate aging and diapause. *Aging Cell* **17**, e12757 (2018).
2. Buffenstein, R. The naked mole-rat: a new long-living model for human aging research. *J. Gerontol. A Biol. Sci. Med. Sci.* **60**, 1369–1377 (2005).
3. Bob Goldstein, N. K. The Future of Cell Biology: Emerging Model Organisms. *Trends Cell Biol.* **26**, 818 (2016).
4. Hochner, B., Shomrat, T. & Fiorito, G. The Octopus: A Model for a Comparative Analysis of the Evolution of Learning and Memory Mechanisms. *The Biological Bulletin* vol. 210 308–317 (2006).
5. Kao, D. *et al.* The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. (2016) doi:10.7554/eLife.20062.
6. The Tabula Microcebus Consortium. Tabula Microcebus: A transcriptomic cell atlas of mouse lemur, an emerging primate model organism. *bioRxiv* 2021.12.12.469460 (2021) doi:10.1101/2021.12.12.469460.
7. Ezran, C. *et al.* The Mouse Lemur, a Genetic Model Organism for Primate Biology, Behavior, and Health. *Genetics* **206**, 651–664 (2017).
8. Ho, C. L. A., Fichtel, C. & Huber, D. The gray mouse lemur (Microcebus murinus) as a model for early primate brain evolution. *Curr. Opin. Neurobiol.* **71**, 92–99 (2021).
9. Languille, S. *et al.* The grey mouse lemur: a non-human primate model for ageing studies. *Ageing Res. Rev.* **11**, 150–162 (2012).
10. Hozer, C., Pifferi, F., Aujard, F. & Perret, M. The Biological Clock in Gray Mouse Lemur: Adaptive, Evolutionary and Aging Considerations in an Emerging Non-human Primate Model. *Front. Physiol.* **10**, 1033 (2019).
11. Perret, M. & Aujard, F. Daily hypothermia and torpor in a tropical primate: synchronization by 24-h light-dark cycle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **281**, (2001).
12. Génin, F., Schilling, A. & Perret, M. Social inhibition of seasonal fattening in wild and captive gray mouse lemurs. *Physiol. Behav.* **86**, 185–194 (2005).
13. Pifferi, F., Epelbaum, J. & Aujard, F. Strengths and Weaknesses of the Gray Mouse Lemur (Microcebus murinus) as a Model for the Behavioral and Psychological Symptoms and Neuropsychiatric Symptoms of Dementia. *Frontiers in Pharmacology* vol. 10 (2019).
14. Marchal, J. *et al.* Calorie restriction and resveratrol supplementation prevent age-related DNA and RNA oxidative damage in a non-human primate. *Exp. Gerontol.* **48**, 992–1000 (2013).
15. Terrien, J. *et al.* Metabolic and genomic adaptations to winter fattening in a primate species, the grey mouse lemur (Microcebus murinus). *International Journal of Obesity* vol. 42 221–230 (2018).
16. Chazarin, B. *et al.* Molecular Liver Fingerprint Reflects the Seasonal Physiology of the Grey Mouse Lemur (Microcebus murinus) during Winter. *Int. J. Mol. Sci.* **23**, (2022).
17. Radespiel, U. Sociality in the gray mouse lemur (Microcebus murinus) in northwestern Madagascar. *Am. J. Primatol.* **51**, (2000).
18. Yoder, A. D. *et al.* Remarkable species diversity in Malagasy mouse lemurs (primates, Microcebus). *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11325–11330 (2000).
19. Rina Evasoa, M. *et al.* Variation in reproduction of the smallest-bodied primate radiation, the mouse lemurs (Microcebus spp.): A synopsis. *Am. J. Primatol.* **80**, e22874 (2018).
20. Hasiniaina, A. F. *et al.* Evolutionary significance of the variation in acoustic communication of a cryptic nocturnal primate radiation (Microcebus spp.). *Ecol. Evol.* **10**, 3784–3797 (2020).
21. Wimmer, B., Tautz, D. & Kappeler, P. The genetic population structure of the gray mouse lemur (Microcebus murinus), a basal primate from Madagascar. *Behavioral Ecology and Sociobiology* vol. 52 166–175 (2002).
22. Larsen, P. A. *et al.* Hybrid de novo genome assembly and centromere characterization of the gray mouse lemur (Microcebus murinus). *BMC Biol.* **15**, 110 (2017).

23. Salzberg, S. L. Next-generation genome annotation: we still struggle to get it right. *Genome Biol.* **20**, 92 (2019).
24. Wang, M. F. Z. *et al.* Uncovering transcriptional dark matter via gene annotation independent single-cell RNA sequencing analysis. *Nat. Commun.* **12**, 2158 (2021).
25. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
26. Botvinnik, O. B. *et al.* Single-cell transcriptomics for the 99.9% of species without reference genomes. *bioRxiv* 2021.07.09.450799 (2021) doi:10.1101/2021.07.09.450799.
27. Zerbino, D. R., Frankish, A. & Flicek, P. Progress, Challenges, and Surprises in Annotating the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **21**, 55–79 (2020).
28. Lefranc, M.-P. *et al.* IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* **43**, D413–22 (2015).
29. Lefranc, M.-P. & Lefranc, G. *The Immunoglobulin FactsBook.* (Academic Press, 2001).
30. Chen, K. & Cerutti, A. The function and regulation of immunoglobulin D. *Curr. Opin. Immunol.* **23**, 345–352 (2011).
31. Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).
32. Dehghannasiri, R., Olivieri, J. E., Damljanovic, A. & Salzman, J. Specific splice junction detection in single cells with SICILIAN. *Genome Biol.* **22**, 219 (2021).
33. Wan, R., Bai, R. & Shi, Y. Molecular choreography of pre-mRNA splicing by the spliceosome. *Curr. Opin. Struct. Biol.* **59**, 124–133 (2019).
34. Pertea, M. *et al.* CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **19**, 208 (2018).
35. Norgren, R. B., Jr. Improving genome assemblies and annotations for nonhuman primates. *ILAR J.* **54**, 144–153 (2013).
36. Tarashansky, A. J., Xue, Y., Li, P., Quake, S. R. & Wang, B. Self-assembling manifolds in single-cell RNA sequencing data. *Elife* **8**, (2019).
37. Tarashansky, A. J. *et al.* Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* vol. 10 (2021).
38. Mallee, J. J., Salvatore, C. A. & LeBourdelles, B. Receptor Activity-modifying Protein 1 Determines the Species Selectivity of Non-peptide CGRP Receptor Antagonists* 210. *Journal of Biological* (2002).
39. Parameswaran, N. & Spielman, W. S. RAMPs: The past, present and future. *Trends Biochem. Sci.* **31**, 631–638 (2006).
40. Liu, S. *et al.* An organism-wide atlas of hormonal signaling based on the mouse lemur single-cell transcriptome. *bioRxiv* (2021) doi:10.1101/2021.12.13.472243.
41. de Groot, N. G., Blokhuis, J. H., Otting, N., Doxiadis, G. G. M. & Bontrop, R. E. Co-evolution of the MHC class I and KIR gene families in rhesus macaques: ancestry and plasticity. *Immunol. Rev.* **267**, 228–245 (2015).
42. Doxiadis, G. G. M. *et al.* Evolutionary stability of MHC class II haplotypes in diverse rhesus macaque populations. *Immunogenetics* **55**, 540–551 (2003).
43. Flajnik, M. F. & Kasahara, M. Comparative genomics of the MHC. *Immunity* **15**, 351–362 (2001).
44. Trowsdale, J. 'Both man & bird & beast': comparative organization of MHC genes. *Immunogenetics* **41**, 1–17 (1995).
45. Wroblewski, E. E., Parham, P. & Guethlein, L. A. Two to Tango: Co-evolution of Hominid Natural Killer Cell Receptors and MHC. *Front. Immunol.* **10**, 177 (2019).
46. Djaoud, Z. & Parham, P. HLAs, TCRs, and KIRs, a Triumvirate of Human Cell-Mediated Immunity. *Annu. Rev. Biochem.* **89**, 717–739 (2020).
47. Norman, P. J. *et al.* Sequences of 95 human MHC haplotypes reveal extreme coding variation in genes other than highly polymorphic HLA class I and II. *Genome Res.* **27**, 813–823 (2017).
48. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules.

Trends Immunol. **37**, 724–737 (2016).
49. Averdam, A. *et al.* Sequence analysis of the grey mouse lemur (Microcebus murinus) MHC class II DQ and DR region. *Immunogenetics* **63**, 85–93 (2011).
50. Guethlein, L. A., Ezran, C., Liu, S., Krasnow, M. A. & Parham, P. Organism-wide mapping of MHC class I and II expression in mouse lemur cells and tissues. *bioRxiv* (2022) doi:10.1101/2022.02.28.482372.
51. Averdam, A. *et al.* A novel system of polymorphic and diverse NK cell receptors in primates. *PLoS Genet.* **5**, e1000688 (2009).
52. Boegel, S. *et al.* HLA and proteasome expression body map. *BMC Med. Genomics* **11**, 36 (2018).
53. Pober, J. S., Merola, J., Liu, R. & Manes, T. D. Antigen Presentation by Vascular Cells. *Front. Immunol.* **8**, (2017).
54. Cerutti, A. *et al.* Regulation of mucosal IgA responses: lessons from primary immunodeficiencies. *Ann. N. Y. Acad. Sci.* **1238**, 132–144 (2011).
55. Collins, A. M., Wang, Y., Roskin, K. M., Marquis, C. P. & Jackson, K. J. L. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, (2015).
56. Boyd, S. D. *et al.* Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J. Immunol.* **184**, 6986–6992 (2010).
57. Popov, A. V., Zou, X., Xian, J., Nicholson, I. C. & Brüggemann, M. A Human Immunoglobulin λ Locus Is Similarly Well Expressed in Mice and Humans. *J. Exp. Med.* **189**, 1611–1620 (1999).
58. Zemlin, M. *et al.* Expressed Murine and Human CDR-H3 Intervals of Equal Length Exhibit Distinct Repertoires that Differ in their Amino Acid Composition and Predicted Range of Structures. *J. Mol. Biol.* **334**, 733–749 (2003).
59. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nature Cell Biology* vol. 22 38–48 (2020).
60. Zhou, B. O., Yue, R., Murphy, M. M., Peyer, J. G. & Morrison, S. J. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. *Cell Stem Cell* **15**, 154–168 (2014).
61. Ambrosi, T. H., Longaker, M. T. & Chan, C. K. F. A Revised Perspective of Skeletal Stem Cell Biology. *Front Cell Dev Biol* **7**, 189 (2019).
62. Capucetti, A., Albano, F. & Bonecchi, R. Multiple Roles for Chemokines in Neutrophil Biology. *Front. Immunol.* **11**, 1259 (2020).
63. Mercier, F. E., Ragu, C. & Scadden, D. T. The bone marrow at the crossroads of blood and immunity. *Nat. Rev. Immunol.* **12**, 49–60 (2011).
64. Ito, T., Carson, W. F., 4th, Cavassani, K. A., Connett, J. M. & Kunkel, S. L. CCR6 as a mediator of immunity in the lung and gut. *Exp. Cell Res.* **317**, 613–619 (2011).
65. Pawelec, P., Ziemka-Nalecz, M., Sypecka, J. & Zalewska, T. The Impact of the CX3CL1/CX3CR1 Axis in Neurological Disorders. *Cells* **9**, (2020).
66. Hickman, S. E., Allison, E. K., Coleman, U., Kingery-Gallagher, N. D. & El Khoury, J. Heterozygous CX3CR1 Deficiency in Microglia Restores Neuronal β-Amyloid Clearance Pathways and Slows Progression of Alzheimer's Like-Disease in PS1-APP Mice. *Front. Immunol.* **0**, (2019).
67. Farnsworth, R. H., Karnezis, T., Maciburko, S. J., Mueller, S. N. & Stacker, S. A. The Interplay Between Lymphatic Vessels and Chemokines. *Front. Immunol.* **10**, 518 (2019).
68. Barone, F. *et al.* Stromal Fibroblasts in Tertiary Lymphoid Structures: A Novel Target in Chronic Inflammation. *Front. Immunol.* **7**, 477 (2016).
69. Bryce, S. A. *et al.* ACKR4 on Stromal Cells Scavenges CCL19 To Enable CCR7-Dependent Trafficking of APCs from Inflamed Skin to Lymph Nodes. *J. Immunol.* **196**, 3341–3353 (2016).
70. Jalkanen, S. & Salmi, M. Lymphatic endothelial cells of the lymph node. *Nat. Rev. Immunol.* **20**, 566–578 (2020).
71. Nie, Y. *et al.* The role of CXCR4 in maintaining peripheral B cell compartments and humoral immunity. *J. Exp. Med.* **200**, 1145–1156 (2004).

72. Fletcher, A. L., Acton, S. E. & Knoblich, K. Lymph node fibroblastic reticular cells in health and disease. *Nat. Rev. Immunol.* **15**, 350–361 (2015).
73. Vermi, W. *et al.* Role of ChemR23 in directing the migration of myeloid and plasmacytoid dendritic cells to lymphoid organs and inflamed skin. *J. Exp. Med.* **201**, 509–515 (2005).
74. Casey, K. M., Karanewsky, C. J., Pendleton, J. L., Krasnow, M. R. & Albertelli, M. A. Fibrous Osteodystrophy, Chronic Renal Disease, and Uterine Adenocarcinoma in Aged Gray Mouse Lemurs (Microcebus murinus). *Comp. Med.* **71**, 256–266 (2021).
75. Metzemaekers, M., Vanheule, V., Janssens, R., Struyf, S. & Proost, P. Overview of the Mechanisms that May Contribute to the Non-Redundant Activities of Interferon-Inducible CXC Chemokine Receptor 3 Ligands. *Front. Immunol.* **8**, 1970 (2017).
76. Kochumon, S. *et al.* Adipose tissue gene expression of CXCL10 and CXCL11 modulates inflammatory markers in obesity: implications for metabolic inflammation and insulin resistance. *Ther. Adv. Endocrinol. Metab.* **11**, 2042018820930902 (2020).
77. Russo, R. C., Garcia, C. C., Teixeira, M. M. & Amaral, F. A. The CXCL8/IL-8 chemokine family and its receptors in inflammatory diseases. *Expert Rev. Clin. Immunol.* **10**, 593–619 (2014).
78. Lawrence, S. M., Corriden, R. & Nizet, V. The Ontogeny of a Neutrophil: Mechanisms of Granulopoiesis and Homeostasis. *Microbiol. Mol. Biol. Rev.* **82**, (2018).
79. Borregaard, N., Sørensen, O. E. & Theilgaard-Mönch, K. Neutrophil granules: a library of innate immunity proteins. *Trends Immunol.* **28**, 340–345 (2007).
80. Hidalgo, A., Chilvers, E. R., Summers, C. & Koenderman, L. The Neutrophil Life Cycle. *Trends Immunol.* **40**, 584–597 (2019).
81. Nomiyama, H., Osada, N. & Yoshie, O. The evolution of mammalian chemokine genes. *Cytokine Growth Factor Rev.* **21**, 253–262 (2010).
82. Fuellen, G., Foell, D., Nacken, W., Sorg, C. & Kerkhoff, C. Absence of S100A12 in mouse: implications for RAGE–S100A12 interaction. *Trends Immunol.* **24**, 622–624 (2003).
83. Ivetic, A., Hoskins Green, H. L. & Hart, S. J. L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling. *Front. Immunol.* **10**, 1068 (2019).
84. Hong, C.-W. Current Understanding in Neutrophil Differentiation and Heterogeneity. *Immune Netw.* **17**, 298–306 (2017).
85. Pan, Z. Z., Parkyn, L., Ray, A. & Ray, P. Inducible lung-specific expression of RANTES: preferential recruitment of neutrophils. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **279**, L658–66 (2000).
86. Metzemaekers, M., Gouwy, M. & Proost, P. Neutrophil chemoattractant receptors in health and disease: double-edged swords. *Cell. Mol. Immunol.* **17**, 433–450 (2020).
87. Adrover, J. M., Nicolás-Ávila, J. A. & Hidalgo, A. Aging: A Temporal Dimension for Neutrophils. *Trends Immunol.* **37**, 334–345 (2016).
88. Casanova-Acebes, M. *et al.* Rhythmic modulation of the hematopoietic niche through neutrophil clearance. *Cell* **153**, 1025–1035 (2013).
89. Shimizu, Y. & Dobashi, K. CC-chemokine CCL15 expression and possible implications for the pathogenesis of IgE-related severe asthma. *Mediators Inflamm.* **2012**, 475253 (2012).
90. Gschwandtner, M., Derler, R. & Midwood, K. S. More Than Just Attractive: How CCL2 Influences Myeloid Cell Behavior Beyond Chemotaxis. *Front. Immunol.* **10**, 2759 (2019).
91. Mendez-Enriquez, E. & García-Zepeda, E. A. The multiple faces of CCL13 in immunity and inflammation. *Inflammopharmacology* **21**, 397–406 (2013).
92. Honda, T., Uehara, T., Matsumoto, G., Arai, S. & Sugano, M. Neutrophil left shift and white blood cell count as markers of bacterial infection. *Clin. Chim. Acta* **457**, 46–53 (2016).
93. Hickman, S. E. *et al.* The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* **16**, 1896–1905 (2013).
94. Jurga, A. M., Paleczna, M. & Kuter, K. Z. Overview of General and Discriminating Markers of Differential Microglia Phenotypes. *Front. Cell. Neurosci.* **14**, 198 (2020).
95. Schyns, J., Bureau, F. & Marichal, T. Lung Interstitial Macrophages: Past, Present, and Future. *J*

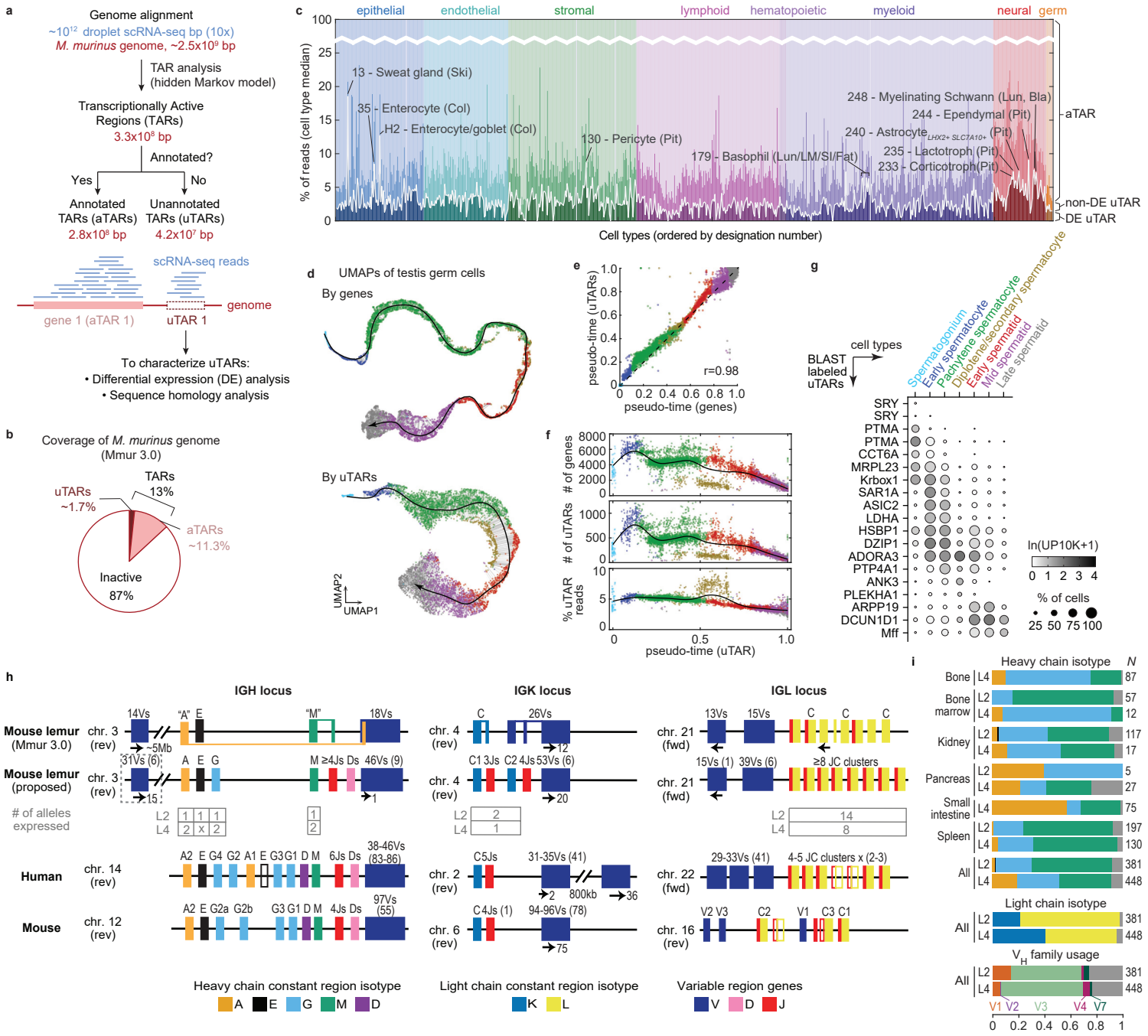
Immunol Res **2018**, 5160794 (2018).
96. Evren, E. *et al.* Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* **54**, 259–275.e7 (2021).
97. Davies, L. C., Jenkins, S. J., Allen, J. E. & Taylor, P. R. Tissue-resident macrophages. *Nat. Immunol.* **14**, 986–995 (2013).
98. Todd, E. M. *et al.* Alveolar macrophage development in mice requires L-plastin for cellular localization in alveoli. *Blood* **128**, 2785–2796 (2016).
99. Borges da Silva, H. *et al.* Splenic Macrophage Subsets and Their Function during Blood-Borne Infections. *Front. Immunol.* **6**, 480 (2015).
100. Kohyama, M. *et al.* Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. *Nature* vol. 457 318–321 (2009).
101. Bonnardel, J. *et al.* Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. *Immunity* **51**, 638–654.e9 (2019).
102. Bain, C. C. & Schridde, A. Origin, Differentiation, and Function of Intestinal Macrophages. *Front. Immunol.* **9**, 2733 (2018).
103. Wu, X. *et al.* Human Liver Macrophage Subsets Defined by CD32. *Front. Immunol.* **11**, 2108 (2020).
104. MacParland, S. A. *et al.* Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
105. Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199–204 (2019).
106. Tsukasaki, M. *et al.* Stepwise cell fate decision pathways during osteoclastogenesis at single-cell resolution. *Nat Metab* **2**, 1382–1390 (2020).
107. Orekhov, A. N. *et al.* Monocyte differentiation and macrophage polarization. *Vessel Plus* **2019**, (2019).
108. Italiani, P. & Boraschi, D. From Monocytes to M1/M2 Macrophages: Phenotypical vs. Functional Differentiation. *Front. Immunol.* **5**, 514 (2014).
109. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* **6**, 13 (2014).
110. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
111. Hu, J. & Sun, J. MUC16 mutations improve patients' prognosis by enhancing the infiltration and antitumor immunity of cytotoxic T lymphocytes in the endometrial cancer microenvironment. *OncoImmunology* vol. 7 e1487914 (2018).
112. Onstad, M. A., Schmandt, R. E. & Lu, K. H. Addressing the Role of Obesity in Endometrial Cancer Risk, Prevention, and Treatment. *J. Clin. Oncol.* **34**, 4225–4230 (2016).
113. Van Nyen, T., Moiola, C. P., Colas, E., Annibali, D. & Amant, F. Modeling Endometrial Cancer: Past, Present, and Future. *Int. J. Mol. Sci.* **19**, (2018).
114. Jurek, B. & Neumann, I. D. The Oxytocin Receptor: From Intracellular Signaling to Behavior. *Physiol. Rev.* **98**, 1805–1908 (2018).
115. Gimpl, G. & Fahrenholz, F. The oxytocin receptor system: structure, function, and regulation. *Physiol. Rev.* **81**, 629–683 (2001).
116. Bochtler, T., Löffler, H. & Krämer, A. Diagnosis and management of metastatic neoplasms with unknown primary. *Semin. Diagn. Pathol.* **35**, 199–206 (2018).
117. Cherchi, P. L. *et al.* The value of serum CA 125 and association CA 125/CA 19-9 in endometrial carcinoma. *Eur. J. Gynaecol. Oncol.* **20**, 315–317 (1999).
118. Moore, R. G. *et al.* Utility of a novel serum tumor biomarker HE4 in patients with endometrioid adenocarcinoma of the uterus. *Gynecol. Oncol.* **110**, 196–201 (2008).
119. Dong, C., Liu, P. & Li, C. Value of HE4 Combined with Cancer Antigen 125 in the Diagnosis of Endometrial Cancer. *Pak. J. Med. Sci. Q.* **33**, 1013–1017 (2017).
120. Cancer Genome Atlas Research Network *et al.* Integrated genomic characterization of endometrial

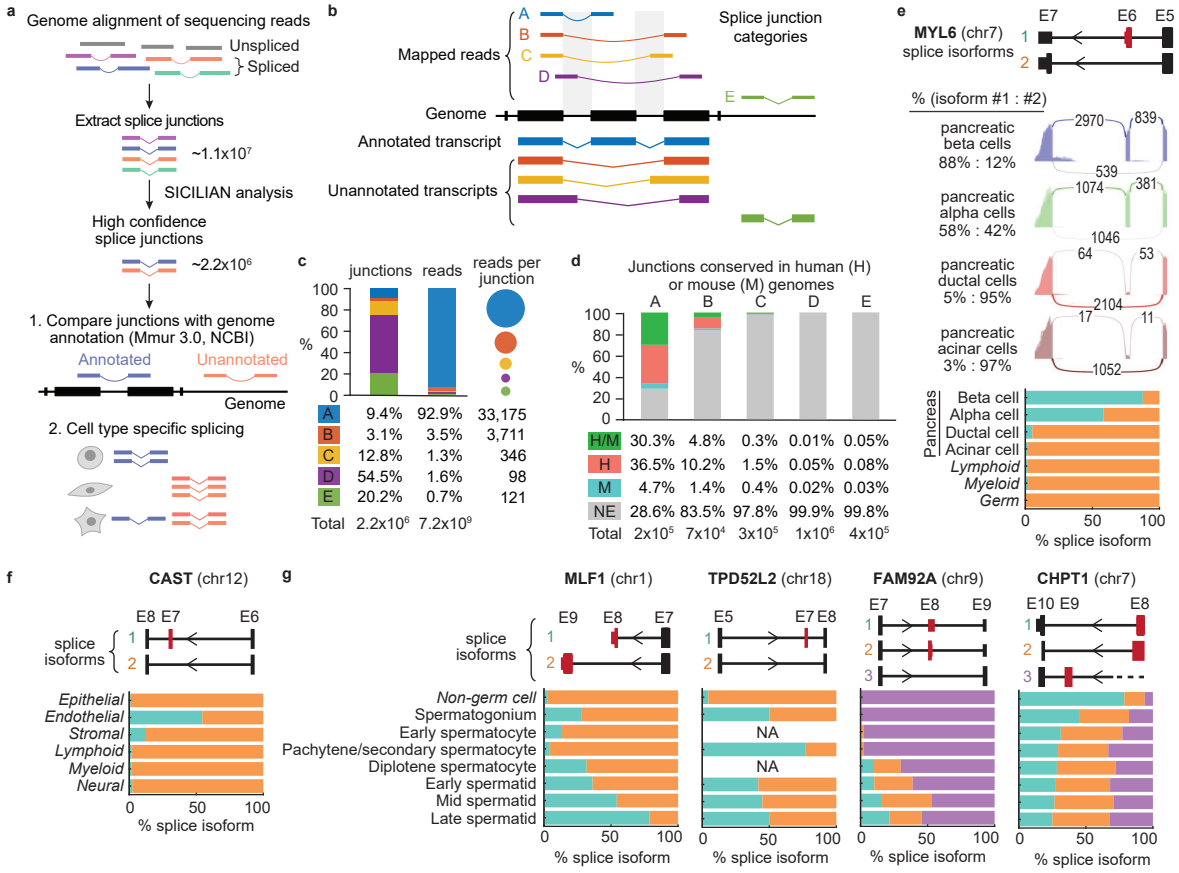
carcinoma. *Nature* **497**, 67–73 (2013).
121. Diver, E. J., Foster, R., Rueda, B. R. & Growdon, W. B. The Therapeutic Challenge of Targeting HER2 in Endometrial Cancer. *Oncologist* **20**, 1058–1068 (2015).
122. Worbs, S. *et al.* Expression of the inhibin/activin subunits (-α, -βA and -βB) in normal and carcinogenic endometrial tissue: Possible immunohistochemical differentiation markers. *Oncol. Rep.* **17**, 97–104 (2007).
123. Xiong, S., Klausen, C., Cheng, J.-C., Zhu, H. & Leung, P. C. K. Activin B induces human endometrial cancer cell adhesion, migration and invasion by up-regulating integrin β3 via SMAD2/3 signaling. *Oncotarget* **6**, 31659–31673 (2015).
124. Backes, F. J. *et al.* Estrogen receptor-alpha as a predictive biomarker in endometrioid endometrial cancer. *Gynecol. Oncol.* **141**, 312–317 (2016).
125. Aghajanian, C. *et al.* Phase II trial of bevacizumab in recurrent or persistent endometrial cancer: a Gynecologic Oncology Group study. *J. Clin. Oncol.* **29**, 2259–2265 (2011).
126. Caretto, M. & Simoncini, T. Progestin or anti-estrogen treatment for endometrial cancer: choosing the best option for selected patients. *Gynecol. Endocrinol.* **37**, 959–960 (2021).
127. van Weelden, W. J., Massuger, L. F. A. G., Enitec, Pijnenborg, J. M. A. & Romano, A. Anti-estrogen Treatment in Endometrial Cancer: A Systematic Review. *Front. Oncol.* **9**, (2019).
128. Perret, M. Stress-effects in Microcebus murinus. *Folia Primatol.* **39**, 63–114 (1982).
129. Terrien, J., Ambid, L., Nibbelink, M., Saint-Charles, A. & Aujard, F. Non-shivering thermogenesis activation and maintenance in the aging gray mouse lemur (Microcebus murinus). *Exp. Gerontol.* **45**, 442–448 (2010).
130. Génin, F., Nibbelink, M., Galand, M., Perret, M. & Ambid, L. Brown fat and nonshivering thermogenesis in the gray mouse lemur (*Microcebus murinus*). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **284**, R811–R818 (2003).
131. Zhang, J. *et al.* Regulation of Torpor in the Gray Mouse Lemur: Transcriptional and Translational Controls and Role of AMPK Signaling. *Genomics, Proteomics & Bioinformatics* vol. 13 103–110 (2015).
132. Ambele, M. A., Dessels, C., Durandt, C. & Pepper, M. S. Genome-wide analysis of gene expression during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene expression during adipocyte differentiation. *Stem Cell Res.* **16**, 725–734 (2016).
133. Menssen, A. *et al.* Differential gene expression profiling of human bone marrow-derived mesenchymal stem cells during adipogenic development. *BMC Genomics* **12**, 461 (2011).
134. Martínez-Sánchez, N. There and Back Again: Leptin Actions in White Adipose Tissue. *Int. J. Mol. Sci.* **21**, (2020).
135. Mustonen, A.-M. *et al.* Circannual leptin and ghrelin levels of the blue fox (Alopex lagopus) in reference to seasonal rhythms of body mass, adiposity, and food intake. *J. Exp. Zool. A Comp. Exp. Biol.* **303**, 26–36 (2005).
136. Ricquier, D. Uncoupling protein 1 of brown adipocytes, the only uncoupler: a historical perspective. *Front. Endocrinol.* **2**, 85 (2011).
137. Chen, Y. *et al.* Crosstalk between KCNK3-Mediated Ion Current and Adrenergic Signaling Regulates Adipose Thermogenesis and Obesity. *Cell* **171**, 836–848.e13 (2017).
138. Jiang, C.-L., Chen, Y.-F. & Lin, F.-J. Apolipoprotein E deficiency activates thermogenesis of white adipose tissues in mice through enhancing β-hydroxybutyrate production from precursor cells. *FASEB J.* **35**, e21760 (2021).
139. Yang, W. *et al.* PPAR γ/Nnat/NF-κB Axis Involved in Promoting Effects of Adiponectin on Preadipocyte Differentiation. *Mediators Inflamm.* **2019**, 5618023 (2019).
140. Petrovic, N. *et al.* Chronic Peroxisome Proliferator-activated Receptor γ (PPARγ) Activation of Epididymally Derived White Adipocyte Cultures Reveals a Population of Thermogenically Competent, UCP1-containing Adipocytes Molecularly Distinct from Classic Brown Adipocytes. *Journal of Biological* (2010).
141. Barneda, D. *et al.* The brown adipocyte protein CIDEA promotes lipid droplet fusion via a

phosphatidic acid-binding amphipathic helix. *Elife* **4**, e07485 (2015).
142. Rosell, M. *et al.* Peroxisome proliferator-activated receptors-α and -γ, and cAMP-mediated pathways, control retinol-binding protein-4 gene expression in brown adipose tissue. *Endocrinology* **153**, 1162–1173 (2012).
143. Szrok, S. *et al.* Metallothioneins 1 and 2, but not 3, are regulated by nutritional status in rat white adipose tissue. *Genes Nutr.* **11**, 1–12 (2016).
144. Do, M.-S. *et al.* Metallothionein gene expression in human adipose tissue from lean and obese subjects. *Horm. Metab. Res.* **34**, 348–351 (2002).
145. Lagathu, C. *et al.* Long-term treatment with interleukin-1beta induces insulin resistance in murine and human adipocytes. *Diabetologia* **49**, 2162–2173 (2006).
146. Shah, R. D. *et al.* Expression of Calgranulin Genes S100A8, S100A9 and S100A12 Is Modulated by n-3 PUFA during Inflammation in Adipose Tissue and Mononuclear Cells. *PLoS One* **12**, e0169614 (2017).
147. Sanger, G. J., Wang, Y., Hobson, A. & Broad, J. Motilin: towards a new understanding of the gastrointestinal neuropharmacology and therapeutic use of motilin receptor agonists. *Br. J. Pharmacol.* **170**, 1323–1332 (2013).
148. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
149. The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* vol. 562 367–372 (2018).
150. Tabula Sapiens Consortium* *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
151. Chae, M., Danko, C. G. & Kraus, W. L. groHMM: a computational tool for identifying unannotated and cell type-specific transcription units from global run-on sequencing data. *BMC Bioinformatics* **16**, 1–16 (2015).
152. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
153. Olivieri, J. E. *et al.* RNA splicing programs define tissue compartments and cell types at single-cell resolution. *Elife* **10**, (2021).
154. Navarro Gonzalez, J. *et al.* The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.* **49**, D1046–D1057 (2021).
155. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
156. Canzar, S., Neu, K. E., Tang, Q., Wilson, P. C. & Khan, A. A. BASIC: BCR assembly from single cells. *Bioinformatics* **33**, 425–427 (2017).
157. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–40 (2013).
158. Flügge, P., Zimmermann, E., Hughes, A. L., Günther, E. & Walter, L. Characterization and phylogenetic relationship of prosimian MHC class I genes. *J. Mol. Evol.* **55**, 768–775 (2002).
159. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
160. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
161. Takeda, A. *et al.* Single-Cell Survey of Human Lymphatics Unveils Marked Endothelial Cell Heterogeneity and Mechanisms of Homing for Neutrophils. *Immunity* **51**, 561–572.e5 (2019).
162. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
163. Ming, J. *et al.* FIRM: Flexible integration of single-cell RNA-sequencing data for large-scale multi-tissue cell atlas datasets. *Brief. Bioinform.* bbac167 (2022).
164. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene

list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* vol. 9 709 (2020).
165. Watson, C. T. *et al.* Complete Haplotype Sequence of the Human Immunoglobulin Heavy-Chain Variable, Diversity, and Joining Genes and Characterization of Allelic and Copy-Number Variation. *Am. J. Hum. Genet.* **92**, 530–546 (2013).
166. Wolf, Y., Yona, S., Kim, K.-W. & Jung, S. Microglia, seen from the CX3CR1 angle. *Front. Cell. Neurosci.* **7**, 26 (2013).
167. Greenbaum, A. M. & Link, D. C. Mechanisms of G-CSF-mediated hematopoietic stem and progenitor mobilization. *Leukemia* **25**, 211–217 (2011).
168. Hartlehnert, M. *et al.* Schwann cells promote post-traumatic nerve inflammation and neuropathic pain through MHC class II. *Sci. Rep.* **7**, 12518 (2017).
169. Cassoni, P. *et al.* Oxytocin receptors in human adenocarcinomas of the endometrium: presence and biological significance. *J. Pathol.* **190**, 470–477 (2000).
170. Halle, M. K. *et al.* HER2 expression patterns in paired primary and metastatic endometrial cancer lesions. *Br. J. Cancer* **118**, 378–387 (2018).
171. Zhang, L. *et al.* Pathogenesis and Clinical Management of Uterine Serous Carcinoma. *Cancers* vol. 12 686 (2020).
172. Jeong, J.-W. *et al.* Mig-6 modulates uterine steroid hormone responsiveness and exhibits altered expression in endometrial disease. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8677–8682 (2009).
173. Holland, C. M., Day, K., Evans, A. & Smith, S. K. Expression of the VEGF and angiopoietin genes in endometrial atypical hyperplasia and endometrial cancer. *Br. J. Cancer* **89**, 891–898 (2003).
174. Yemelyanova, A. *et al.* PAX8 expression in uterine adenocarcinomas and mesonephric proliferations. *Int. J. Gynecol. Pathol.* **33**, 492–499 (2014).
175. Le Gallo, M. & Bell, D. W. The emerging genomic landscape of endometrial cancer. *Clin. Chem.* **60**, 98–110 (2014).
176. Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**, 683–695 (2005).
177. Maccari, G. *et al.* IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Res.* **45**, D860–D864 (2017).
178. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **45**, D12–D17 (2017).
179. Pham, T. D. *et al.* High-fat diet induces systemic B-cell repertoire changes associated with insulin resistance. *Mucosal Immunol.* **10**, 1468–1479 (2017).
180. Sankar, K., Hoi, K. H. & Hötzel, I. Dynamics of heavy chain junctional length biases in antibody repertoires. *Commun Biol* **3**, 207 (2020).
181. Skovdahl, H. K. *et al.* Expression of CCL20 and Its Corresponding Receptor CCR6 Is Enhanced in Active Inflammatory Bowel Disease, and TLR3 Mediates CCL20 Expression in Colonic Epithelial Cells. *PLOS ONE* vol. 10 e0141710 (2015).
182. Sampath, P., Moideen, K., Ranganathan, U. D. & Bethunaickan, R. Monocyte Subsets: Phenotypes and Function in Tuberculosis Infection. *Front. Immunol.* **9**, 1726 (2018).
183. Shi, C. & Pamer, E. G. Monocyte recruitment during infection and inflammation. *Nat. Rev. Immunol.* **11**, 762–774 (2011).

Fig. 1


a
Genome alignment
~10^12 droplet scRNA-seq bp (10x)
M. murinus genome, ~2.5x10^9 bp
TAR analysis (hidden Markov model)
Transcriptionally Active Regions (TARs)
3.3x10^8 bp
Annotated?
Yes
No
Annotated TARs (aTARs)
2.8x10^8 bp
Unannotated TARs (uTARs)
4.2x10^7 bp
scRNA-seq reads
genome
gene 1 (aTAR 1)
uTAR 1
To characterize uTARs:
• Differential expression (DE) analysis
• Sequence homology analysis
b
Coverage of M. murinus genome (Mmur 3.0)
TARs 13%
uTARs ~1.7%
aTARs ~11.3%
Inactive 87%
c
epithelial
endothelial
stromal
lymphoid
hematopoietic
myeloid
neural
germ
% of reads (cell type median)
13 - Sweat gland (Ski)
35 - Enterocyte (Col)
H2 - Enterocyte/goblet (Col)
130 - Pericyte (Pit)
179 - Basophil (Lun/LM/SI/Fat)
248 - Myelinating Schwann (Lun, Bla)
244 - Ependymal (Pit)
240 - Astrocyte LHX2+ SLC7A10+ (Pit)
235 - Lactotroph (Pit)
233 - Corticotroph (Pit)
aTAR
non-DE uTAR
DE uTAR
Cell types (ordered by designation number)
d
UMAPs of testis germ cells
By genes
By uTARs
UMAP2
UMAP1
e
pseudo-time (uTARs)
r=0.98
pseudo-time (genes)
f
# of genes
# of uTARs
% uTAR reads
pseudo-time (uTAR)
g
cell types
BLAST labeled uTARs
Spermatogonium
Early spermatocyte
Pachytene spermatocyte
Diplotene/secondary spermatocyte
Early spermatid
Mid spermatid
Late spermatid
SRY
PTMA
CCT6A
MRPL23
Krbox1
SAR1A
ASIC2
LDHA
HSBP1
DZIP1
ADORA3
PTP4A1
ANK3
PLEKHA1
ARPP19
DCUN1D1
Mff
ln(UP10K+1)
% of cells
h
IGH locus
IGK locus
IGL locus
Mouse lemur (Mmur 3.0)
Mouse lemur (proposed)
# of alleles expressed
Human
Mouse
Heavy chain constant region isotype
Light chain constant region isotype
Variable region genes
i
Heavy chain isotype
N
Bone
Bone marrow
Kidney
Pancreas
Small intestine
Spleen
All
Light chain isotype
V_H family usage


a
Genome alignment of sequencing reads
Unspliced
Spliced
Extract splice junctions
~1.1x10^7
SICILIAN analysis
High confidence splice junctions
~2.2x10^6
1. Compare junctions with genome annotation (Mmur 3.0, NCBI)
Annotated
Unannotated
Genome
2. Cell type specific splicing
b
Mapped reads
A
B
C
D
E
Splice junction categories
Genome
Annotated transcript
Unannotated transcripts
c
junctions
reads
reads per junction
%
A 9.4% 92.9% 33,175
B 3.1% 3.5% 3,711
C 12.8% 1.3% 346
D 54.5% 1.6% 98
E 20.2% 0.7% 121
Total 2.2x10^6 7.2x10^9
d
Junctions conserved in human (H) or mouse (M) genomes
A B C D E
%
H/M 30.3% 4.8% 0.3% 0.01% 0.05%
H 36.5% 10.2% 1.5% 0.05% 0.08%
M 4.7% 1.4% 0.4% 0.02% 0.03%
NE 28.6% 83.5% 97.8% 99.9% 99.8%
Total 2x10^5 7x10^4 3x10^5 1x10^6 4x10^5
e
MYL6 (chr7)
splice isoforms
E7 E6 E5
% (isoform #1 : #2)
pancreatic beta cells 88% : 12%
2970 839 539
pancreatic alpha cells 58% : 42%
1074 381 1046
pancreatic ductal cells 5% : 95%
64 53 2104
pancreatic acinar cells 3% : 97%
17 11 1052
Pancreas
Beta cell
Alpha cell
Ductal cell
Acinar cell
Lymphoid
Myeloid
Germ
% splice isoform
f
CAST (chr12)
E8 E7 E6
splice isoforms
Epithelial
Endothelial
Stromal
Lymphoid
Myeloid
Neural
% splice isoform
g
MLF1 (chr1)
E9 E8 E7
TPD52L2 (chr18)
E5 E7 E8
FAM92A (chr9)
E7 E8 E9
CHPT1 (chr7)
E10 E9 E8
splice isoforms
Non-germ cell
Spermatogonium
Early spermatocyte
Pachytene/secondary spermatocyte
Diplotene spermatocyte
Early spermatid
Mid spermatid
Late spermatid
NA
NA
% splice isoform

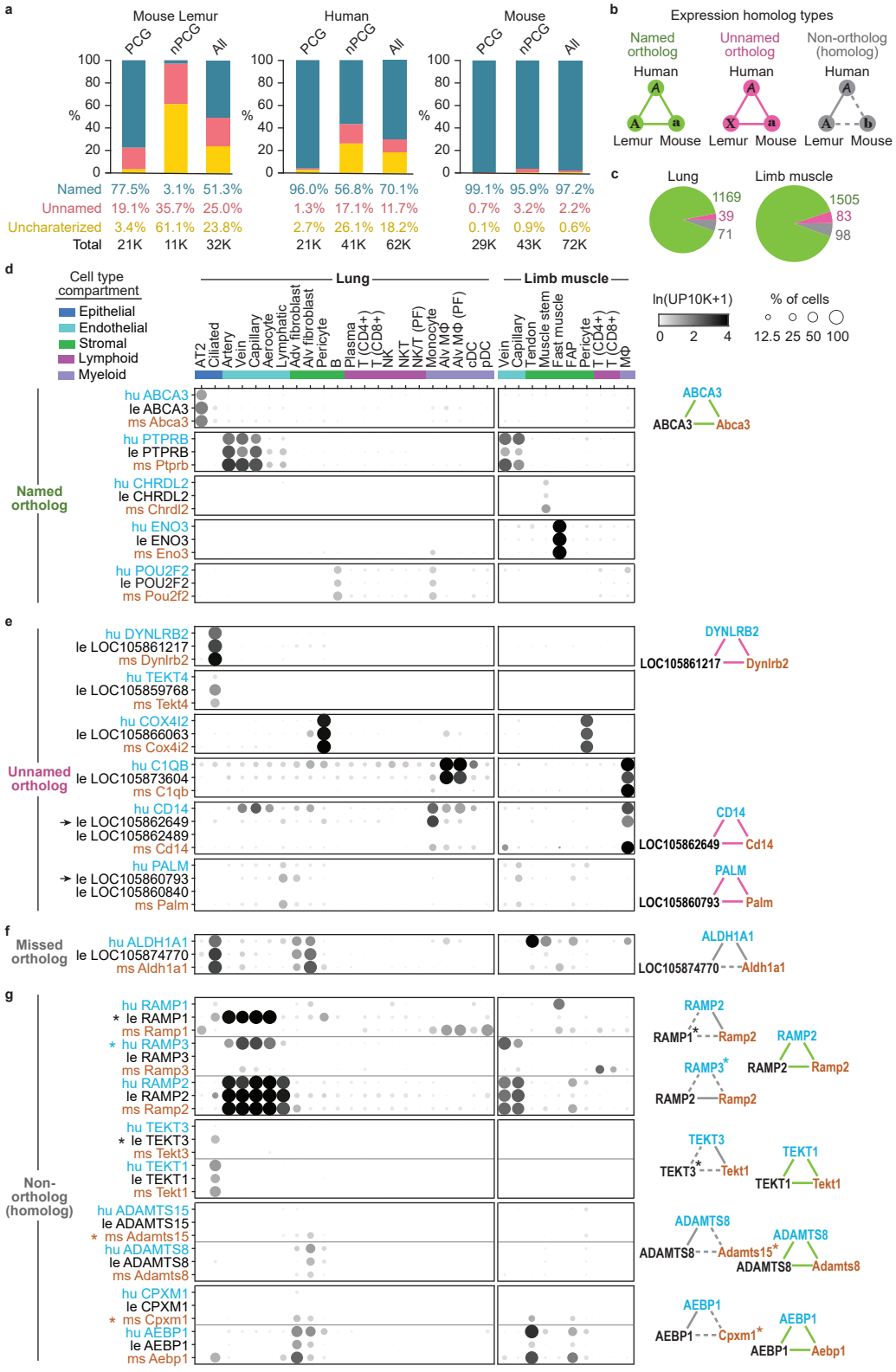

a
Mouse Lemur
Human
Mouse
PCG
nPCG
All
%
Named 77.5% 3.1% 51.3% 96.0% 56.8% 70.1% 99.1% 95.9% 97.2%
Unnamed 19.1% 35.7% 25.0% 1.3% 17.1% 11.7% 0.7% 3.2% 2.2%
Uncharaterized 3.4% 61.1% 23.8% 2.7% 26.1% 18.2% 0.1% 0.9% 0.6%
Total 21K 11K 32K 21K 41K 62K 29K 43K 72K
b
Expression homolog types
Named ortholog
Unnamed ortholog
Non-ortholog (homolog)
Human
Lemur Mouse
c
Lung
Limb muscle
1169
39
71
1505
83
98
d
Cell type compartment
Epithelial
Endothelial
Stromal
Lymphoid
Myeloid
Lung
Limb muscle
AT2
Ciliated
Artery
Vein
Capillary
Aerocyte
Lymphatic
Adv fibroblast
Alv fibroblast
Pericyte
B
Plasma
T (CD4+)
T (CD8+)
NK
NKT
NK/T (PF)
Monocyte
Alv MΦ
Alv MΦ (PF)
cDC
pDC
Vein
Capillary
Tendon
Muscle stem
Fast muscle
FAP
Pericyte
T (CD4+)
T (CD8+)
MΦ
ln(UP10K+1)
0 2 4
% of cells
12.5 25 50 100
Named ortholog
hu ABCA3
le ABCA3
ms Abca3
hu PTPRB
le PTPRB
ms Ptprb
hu CHRDL2
le CHRDL2
ms Chrdl2
hu ENO3
le ENO3
ms Eno3
hu POU2F2
le POU2F2
ms Pou2f2
ABCA3
ABCA3
Abca3
e
Unnamed ortholog
hu DYNLRB2
le LOC105861217
ms Dynlrb2
hu TEKT4
le LOC105859768
ms Tekt4
hu COX4I2
le LOC105866063
ms Cox4i2
hu C1QB
le LOC105873604
ms C1qb
hu CD14
le LOC105862649
le LOC105862489
ms Cd14
hu PALM
le LOC105860793
le LOC105860840
ms Palm
DYNLRB2
LOC105861217
Dynlrb2
CD14
LOC105862649
Cd14
PALM
LOC105860793
Palm
f
Missed ortholog
hu ALDH1A1
le LOC105874770
ms Aldh1a1
ALDH1A1
LOC105874770
Aldh1a1
g
Non-ortholog (homolog)
hu RAMP1
le RAMP1
ms Ramp1
hu RAMP3
le RAMP3
ms Ramp3
hu RAMP2
le RAMP2
ms Ramp2
hu TEKT3
le TEKT3
ms Tekt3
hu TEKT1
le TEKT1
ms Tekt1
hu ADAMTS15
le ADAMTS15
ms Adamts15
hu ADAMTS8
le ADAMTS8
ms Adamts8
hu CPXM1
le CPXM1
ms Cpxm1
hu AEBP1
le AEBP1
ms Aebp1
RAMP2
RAMP1
Ramp2
RAMP3
TEKT3
Tekt1
TEKT1
ADAMTS8
Adamts15
Adamts8
AEBP1
Cpxm1
Aebp1

Fig. 4

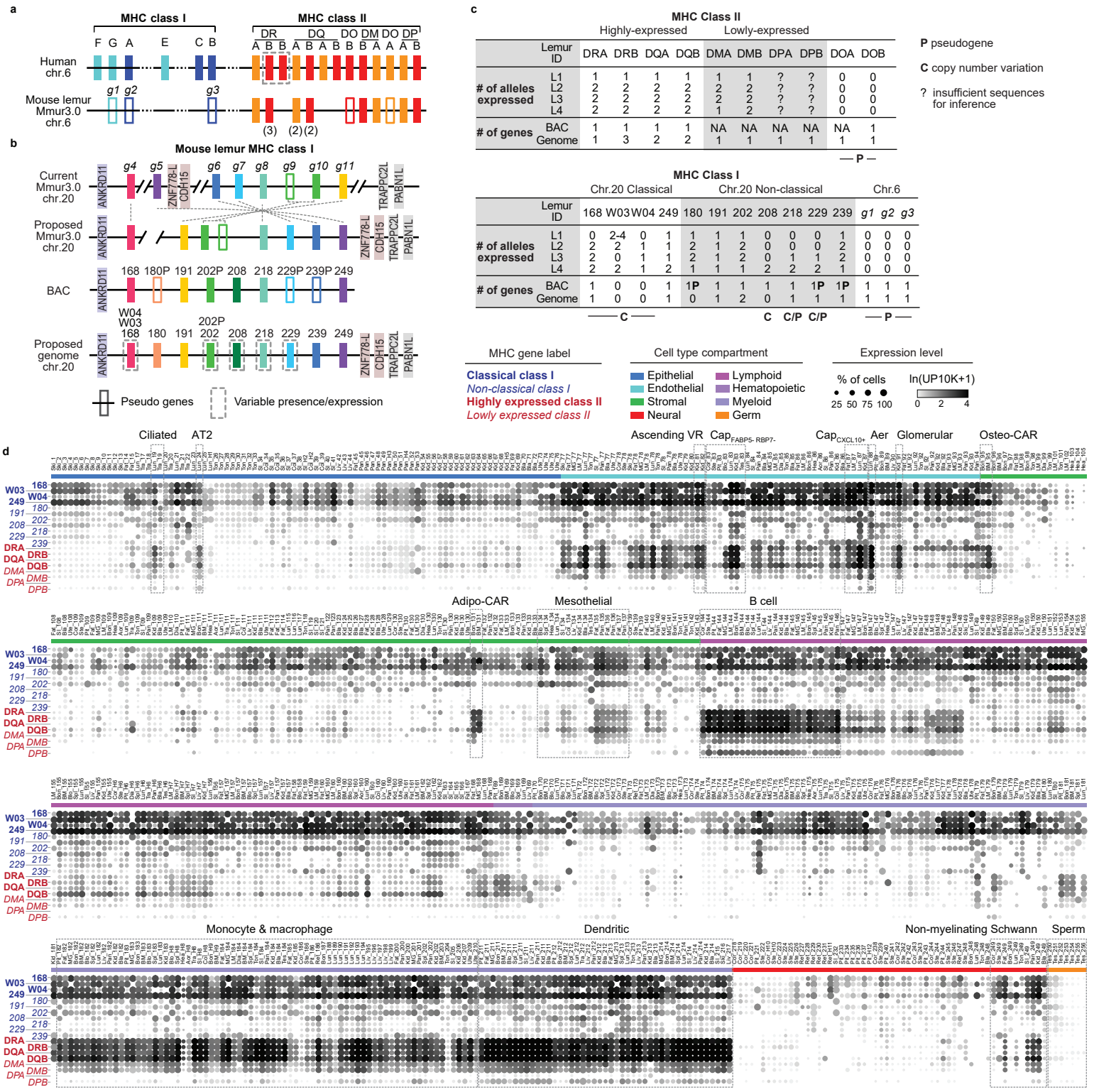
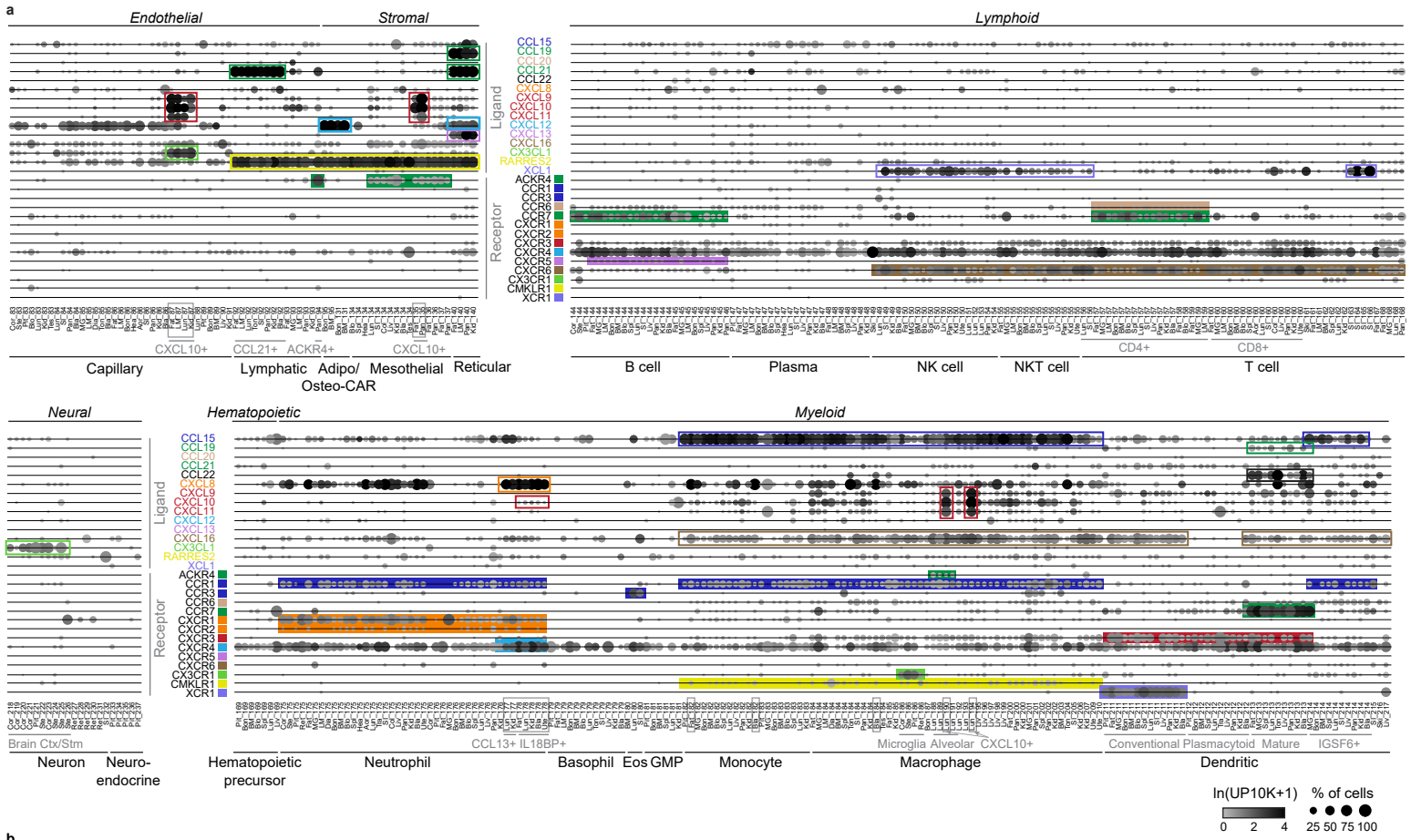
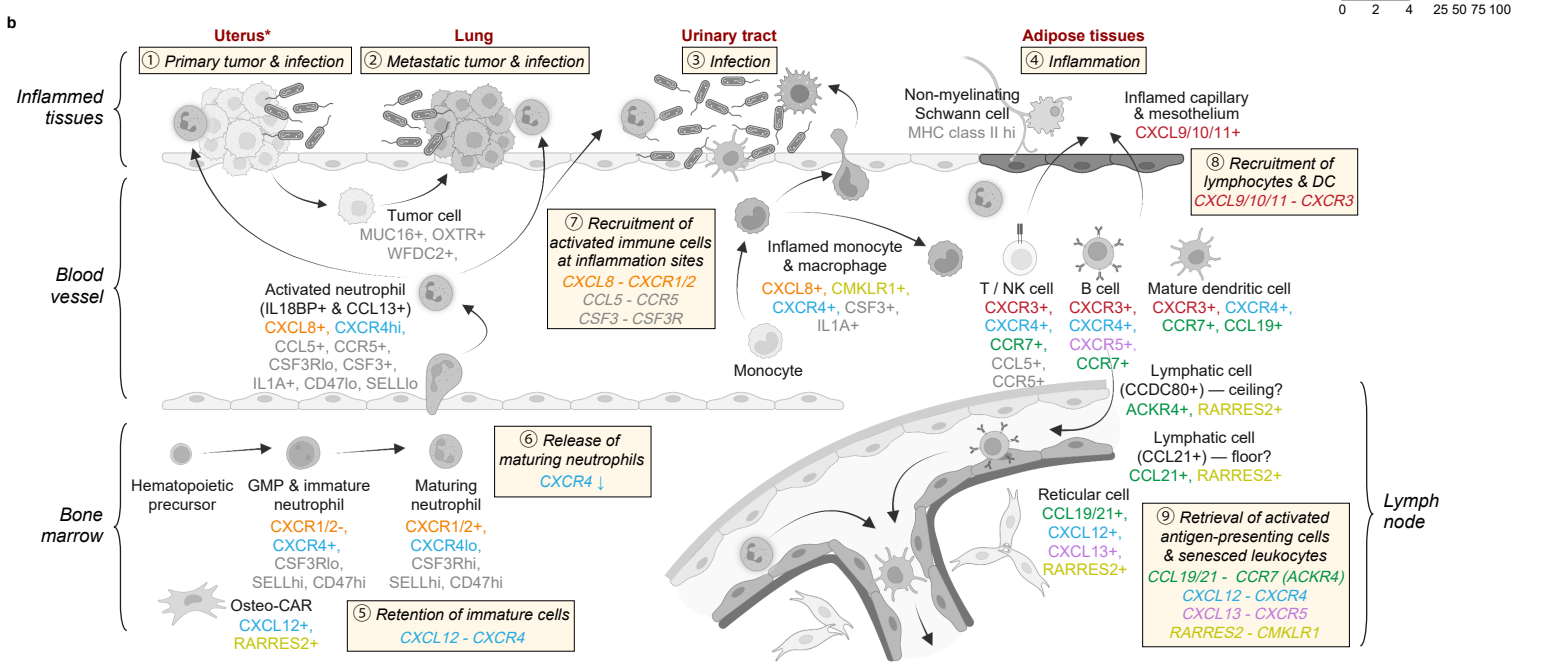
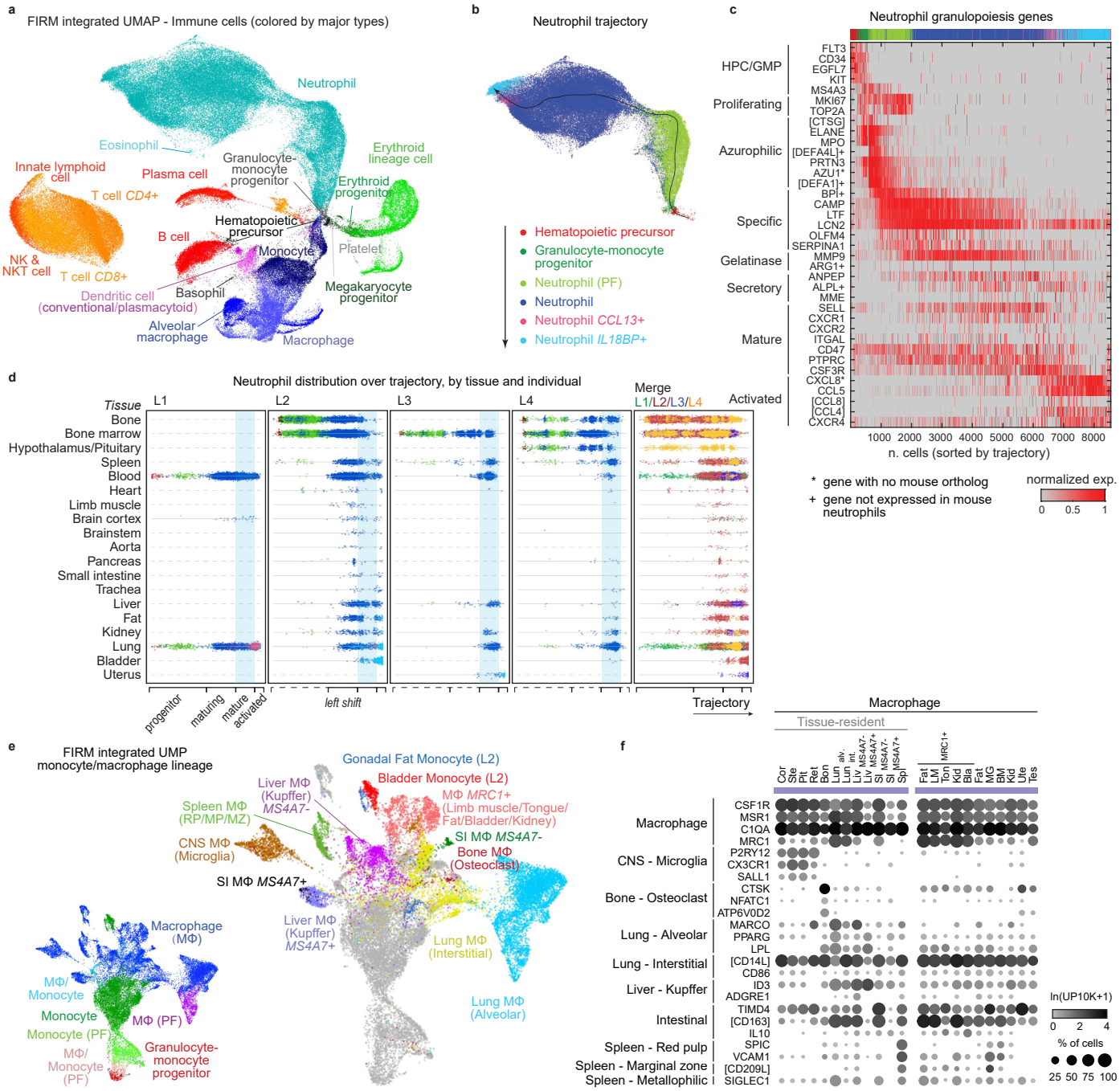

a
MHC class I
MHC class II
F G A E C B
DR DQ DO DM DO DP
A B B A B A B B B A A A B
Human chr.6
Mouse lemur Mmur3.0 chr.6
g1 g2 g3
(3) (2)(2)
b
Mouse lemur MHC class I
Current Mmur3.0 chr.20
Proposed Mmur3.0 chr.20
BAC
Proposed genome chr.20
ANKRD11
ZNF778-L
CDH15
TRAPPC2L
PABN1L
g4 g5 g6 g7 g8 g9 g10 g11
168 180P 191 202P 208 218 229P 239P 249
W04 W03 202P
168 180 191 202 208 218 229 239 249
Pseudo genes
Variable presence/expression
c
MHC Class II
Highly-expressed
Lowly-expressed
Lemur ID
DRA DRB DQA DQB DMA DMB DPA DPB DOA DOB
# of alleles expressed
L1 1 1 1 1 1 1 ? ? 0 0
L2 2 2 2 2 2 2 ? ? 0 0
L3 2 2 2 2 2 2 ? ? 0 0
L4 2 2 2 2 1 2 ? ? 0 0
# of genes
BAC 1 1 1 1 NA NA NA NA NA 1
Genome 1 3 2 2 1 1 1 1 1 1
P
P pseudogene
C copy number variation
? insufficient sequences for inference
MHC Class I
Chr.20 Classical
Chr.20 Non-classical
Chr.6
Lemur ID
168 W03 W04 249 180 191 202 208 218 229 239 g1 g2 g3
# of alleles expressed
L1 0 2-4 0 1 1 1 1 0 0 0 1 0 0 0
L2 2 2 1 1 2 1 2 0 0 0 2 0 0 0
L3 2 0 1 1 2 1 2 0 1 1 2 0 0 0
L4 2 2 1 2 1 1 1 2 2 2 1 0 0 0
# of genes
BAC 1 0 0 1 1P 1 1 1 1 1P 1P 1 1 1
Genome 1 0 0 1 0 1 2 0 1 1 1 1 1 1
C
C C/P C/P
P
MHC gene label
Classical class I
Non-classical class I
Highly expressed class II
Lowly expressed class II
Cell type compartment
Epithelial
Endothelial
Stromal
Neural
Lymphoid
Hematopoietic
Myeloid
Germ
Expression level
% of cells
25 50 75 100
ln(UP10K+1)
0 2 4
d
Ciliated
AT2
Ascending VR
Cap FABP5- RBP7-
Cap CXCL10+
Aer
Glomerular
Osteo-CAR
Adipo-CAR
Mesothelial
B cell
Monocyte & macrophage
Dendritic
Non-myelinating Schwann
Sperm
W03 W04 249 168 191 180 202 208 218 229 239 DRA DRB DQA DQB DMA DMB DPA DPB

Fig 5


a
Endothelial
Stromal
Lymphoid
Ligand
Receptor
CCL15
CCL19
CCL20
CCL21
CCL22
CXCL8
CXCL9
CXCL10
CXCL11
CXCL12
CXCL13
CXCL16
CX3CL1
RARRES2
XCL1
ACKR4
CCR1
CCR3
CCR6
CCR7
CXCR1
CXCR2
CXCR3
CXCR4
CXCR5
CXCR6
CX3CR1
CMKLR1
XCR1
CXCL10+
CCL21+
ACKR4+
CXCL10+
Capillary
Lymphatic
Adipo/ Osteo-CAR
Mesothelial
Reticular
B cell
Plasma
NK cell
NKT cell
CD4+
CD8+
T cell
Neural
Hematopoietic
Myeloid
Brain Ctx/Stm
CCL13+ IL18BP+
Microglia Alveolar CXCL10+
Conventional Plasmacytoid Mature IGSF6+
Neuron
Neuro-endocrine
Hematopoietic precursor
Neutrophil
Basophil Eos GMP
Monocyte
Macrophage
Dendritic
ln(UP10K+1)
0 2 4
% of cells
25 50 75 100


b
Uterus*
Lung
Urinary tract
Adipose tissues
Inflammed tissues
① Primary tumor & infection
② Metastatic tumor & infection
③ Infection
④ Inflammation
Non-myelinating Schwann cell
MHC class II hi
Inflamed capillary & mesothelium
CXCL9/10/11+
⑧ Recruitment of lymphocytes & DC
CXCL9/10/11 - CXCR3
Blood vessel
Tumor cell
MUC16+, OXTR+
WFDC2+,
⑦ Recruitment of activated immune cells at inflammation sites
CXCL8 - CXCR1/2
CCL5 - CCR5
CSF3 - CSF3R
Inflamed monocyte & macrophage
CXCL8+, CMKLR1+,
CXCR4+, CSF3+,
IL1A+
Activated neutrophil
(IL18BP+ & CCL13+)
CXCL8+, CXCR4hi,
CCL5+, CCR5+,
CSF3Rlo, CSF3+,
IL1A+, CD47lo, SELLlo
Monocyte
T / NK cell
CXCR3+,
CXCR4+,
CCR7+,
CCL5+,
CCR5+
B cell
CXCR3+,
CXCR4+,
CXCR5+,
CCR7+
Mature dendritic cell
CXCR3+, CXCR4+,
CCR7+, CCL19+
Lymphatic cell
(CCDC80+) — ceiling?
ACKR4+, RARRES2+
Lymphatic cell
(CCL21+) — floor?
CCL21+, RARRES2+
Bone marrow
Hematopoietic precursor
GMP & immature neutrophil
CXCR1/2-,
CXCR4+,
CSF3Rlo,
SELLhi, CD47hi
Maturing neutrophil
CXCR1/2+,
CXCR4lo,
CSF3Rhi,
SELLhi, CD47hi
⑥ Release of maturing neutrophils
CXCR4 ↓
Osteo-CAR
CXCL12+,
RARRES2+
⑤ Retention of immature cells
CXCL12 - CXCR4
Reticular cell
CCL19/21+,
CXCL12+,
CXCL13+,
RARRES2+
⑨ Retrieval of activated antigen-presenting cells & senesced leukocytes
CCL19/21 - CCR7 (ACKR4)
CXCL12 - CXCR4
CXCL13 - CXCR5
RARRES2 - CMKLR1
Lymph node


a
FIRM integrated UMAP - Immune cells (colored by major types)
Neutrophil
Eosinophil
Innate lymphoid cell
Plasma cell
Granulocyte-monocyte progenitor
Erythroid lineage cell
T cell CD4+
Erythroid progenitor
Hematopoietic precursor
B cell
Monocyte
Platelet
NK & NKT cell
T cell CD8+
Dendritic cell (conventional/plasmacytoid)
Basophil
Megakaryocyte progenitor
Alveolar macrophage
Macrophage
b
Neutrophil trajectory
Hematopoietic precursor
Granulocyte-monocyte progenitor
Neutrophil (PF)
Neutrophil
Neutrophil CCL13+
Neutrophil IL18BP+
c
Neutrophil granulopoiesis genes
HPC/GMP
Proliferating
Azurophilic
Specific
Gelatinase
Secretory
Mature
Activated
FLT3
CD34
EGFL7
KIT
MS4A3
MKI67
TOP2A
[CTSG]
ELANE
MPO
[DEFA4L]+
PRTN3
AZU1*
[DEFA1]+
BPI+
CAMP
LTF
LCN2
OLFM4
SERPINA1
MMP9
ARG1+
ANPEP
ALPL+
MME
SELL
CXCR1
CXCR2
ITGAL
CD47
PTPRC
CSF3R
CXCL8*
CCL5
[CCL8]
[CCL4]
CXCR4
1000 2000 3000 4000 5000 6000 7000 8000
n. cells (sorted by trajectory)
* gene with no mouse ortholog
+ gene not expressed in mouse neutrophils
normalized exp.
0 0.5 1
d
Neutrophil distribution over trajectory, by tissue and individual
Tissue
L1
L2
L3
L4
Merge L1/L2/L3/L4
Bone
Bone marrow
Hypothalamus/Pituitary
Spleen
Blood
Heart
Limb muscle
Brain cortex
Brainstem
Aorta
Pancreas
Small intestine
Trachea
Liver
Fat
Kidney
Lung
Bladder
Uterus
progenitor
maturing
mature
activated
left shift
Trajectory
e
FIRM integrated UMP monocyte/macrophage lineage
Gonadal Fat Monocyte (L2)
Bladder Monocyte (L2)
Liver MΦ (Kupffer) MS4A7-
MΦ MRC1+ (Limb muscle/Tongue/Fat/Bladder/Kidney)
Spleen MΦ (RP/MP/MZ)
SI MΦ MS4A7-
CNS MΦ (Microglia)
Bone MΦ (Osteoclast)
SI MΦ MS4A7+
Macrophage (MΦ)
Liver MΦ (Kupffer) MS4A7+
Lung MΦ (Interstitial)
MΦ/ Monocyte
Monocyte
MΦ (PF)
Monocyte (PF)
Lung MΦ (Alveolar)
MΦ/ Monocyte (PF)
Granulocyte-monocyte progenitor
f
Macrophage
Tissue-resident
MRC1+
Cor Ste Plt Ret Bon Lun Lun int. Liv alv. Liv MS4A7- SI MS4A7+ SI MS4A7- Spl MS4A7+
Fat LM Ton Kid Bla Fat MG BM Kid Ute Tes
Macrophage
CSF1R
MSR1
C1QA
MRC1
CNS - Microglia
P2RY12
CX3CR1
SALL1
Bone - Osteoclast
CTSK
NFATC1
ATP6V0D2
Lung - Alveolar
MARCO
PPARG
LPL
Lung - Interstitial
[CD14L]
CD86
Liver - Kupffer
ID3
ADGRE1
Intestinal
TIMD4
[CD163]
IL10
Spleen - Red pulp
SPIC
Spleen - Marginal zone
VCAM1
[CD209L]
Spleen - Metallophilic
SIGLEC1
ln(UP10K+1)
0 2 4
% of cells
25 50 75 100

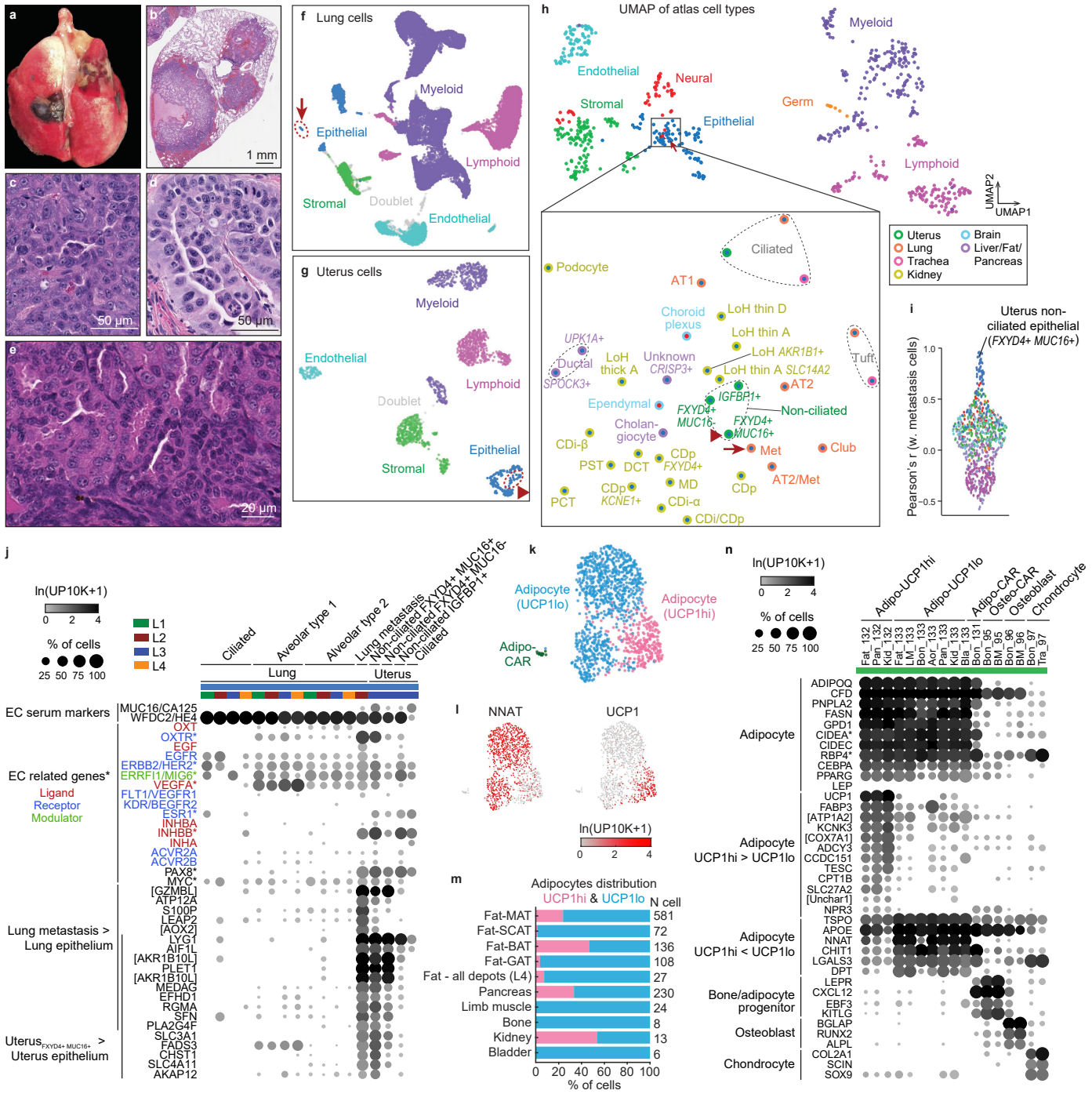

a
b
1 mm
c
50 µm
d
50 µm
e
20 µm
f Lung cells
Myeloid
Epithelial
Lymphoid
Stromal
Doublet
Endothelial
g Uterus cells
Myeloid
Endothelial
Lymphoid
Doublet
Epithelial
Stromal
h
UMAP of atlas cell types
Endothelial
Neural
Stromal
Epithelial
Germ
Myeloid
Lymphoid
UMAP2
UMAP1
Uterus
Lung
Trachea
Kidney
Brain
Liver/Fat/ Pancreas
Podocyte
Ciliated
AT1
Choroid plexus
LoH thin D
LoH thin A
UPK1A+
Ductal
SPOCK3+
LoH thick A
Unknown
CRISP3+
LoH AKR1B1+
LoH thin A SLC14A2
Tuft
AT2
IGFBP1+
Ependymal
FXYD4+ MUC16-
Non-ciliated
Cholan-giocyte
FXYD4+ MUC16+
CDi-β
Met
Club
PST
DCT
CDp FXYD4+
MD
CDp
AT2/Met
PCT
CDp KCNE1+
CDi-α
CDi/CDp
i
Uterus non-ciliated epithelial (FXYD4+ MUC16+)
Pearson's r (w. metastasis cells)
1.0
0.5
0.0
−0.5
j
ln(UP10K+1)
0 2 4
% of cells
25 50 75 100
L1
L2
L3
L4
Ciliated
Aveolar type 1
Aveolar type 2
Lung metastasis
Non-ciliated FXYD4+ MUC16+
Non-ciliated FXYD4+ MUC16-
Non-ciliated IGFBP1+
Ciliated
Lung
Uterus
EC serum markers
MUC16/CA125
WFDC2/HE4
EC related genes*
Ligand
Receptor
Modulator
OXT
OXTR*
EGF
EGFR
ERBB2/HER2*
ERRFI1/MIG6*
VEGFA*
FLT1/VEGFR1
KDR/BEGFR2
ESR1*
INHBA
INHBB*
INHA
ACVR2A
ACVR2B
PAX8*
MYC*
Lung metastasis > Lung epithelium
[GZMBL]
ATP12A
S100P
LEAP2
[AOX2]
LYG1
AIF1L
[AKR1B10L]
PLET1
[AKR1B10L]
MEDAG
EFHD1
RGMA
SFN
Uterus FXYD4+ MUC16+ > Uterus epithelium
PLA2G4F
SLC3A1
FADS3
CHST1
SLC4A11
AKAP12
k
Adipocyte (UCP1lo)
Adipocyte (UCP1hi)
Adipo-CAR
l
NNAT
UCP1
ln(UP10K+1)
0 2 4
m
Adipocytes distribution
UCP1hi & UCP1lo
N cell
Fat-MAT 581
Fat-SCAT 72
Fat-BAT 136
Fat-GAT 108
Fat - all depots (L4) 27
Pancreas 230
Limb muscle 24
Bone 8
Kidney 13
Bladder 6
0 20 40 60 80 100
% of cells
n
ln(UP10K+1)
0 2 4
% of cells
25 50 75 100
Adipo-UCP1hi
Adipo-UCP1lo
Adipo-CAR
Osteo-CAR
Osteoblast
Chondrocyte
Fat_132
Pan_132
Kid_132
Fat_133
LM_133
Bon_133
Aor_133
Pan_133
Kid_133
Bla_133
Bon_131
Bon_95
BM_95
Bon_96
BM_96
Bon_97
Tra_97
Adipocyte
ADIPOQ
CFD
PNPLA2
FASN
GPD1
CIDEA*
CIDEC
RBP4*
CEBPA
PPARG
LEP
Adipocyte UCP1hi > UCP1lo
UCP1
FABP3
[ATP1A2]
KCNK3
[COX7A1]
ADCY3
CCDC151
TESC
CPT1B
SLC27A2
[Unchar1]
NPR3
Adipocyte UCP1hi < UCP1lo
TSPO
APOE
NNAT
CHIT1
LGALS3
DPT
Bone/adipocyte progenitor
LEPR
CXCL12
EBF3
KITLG
Osteoblast
BGLAP
RUNX2
ALPL
Chondrocyte
COL2A1
SCIN
SOX9

Fig. 8

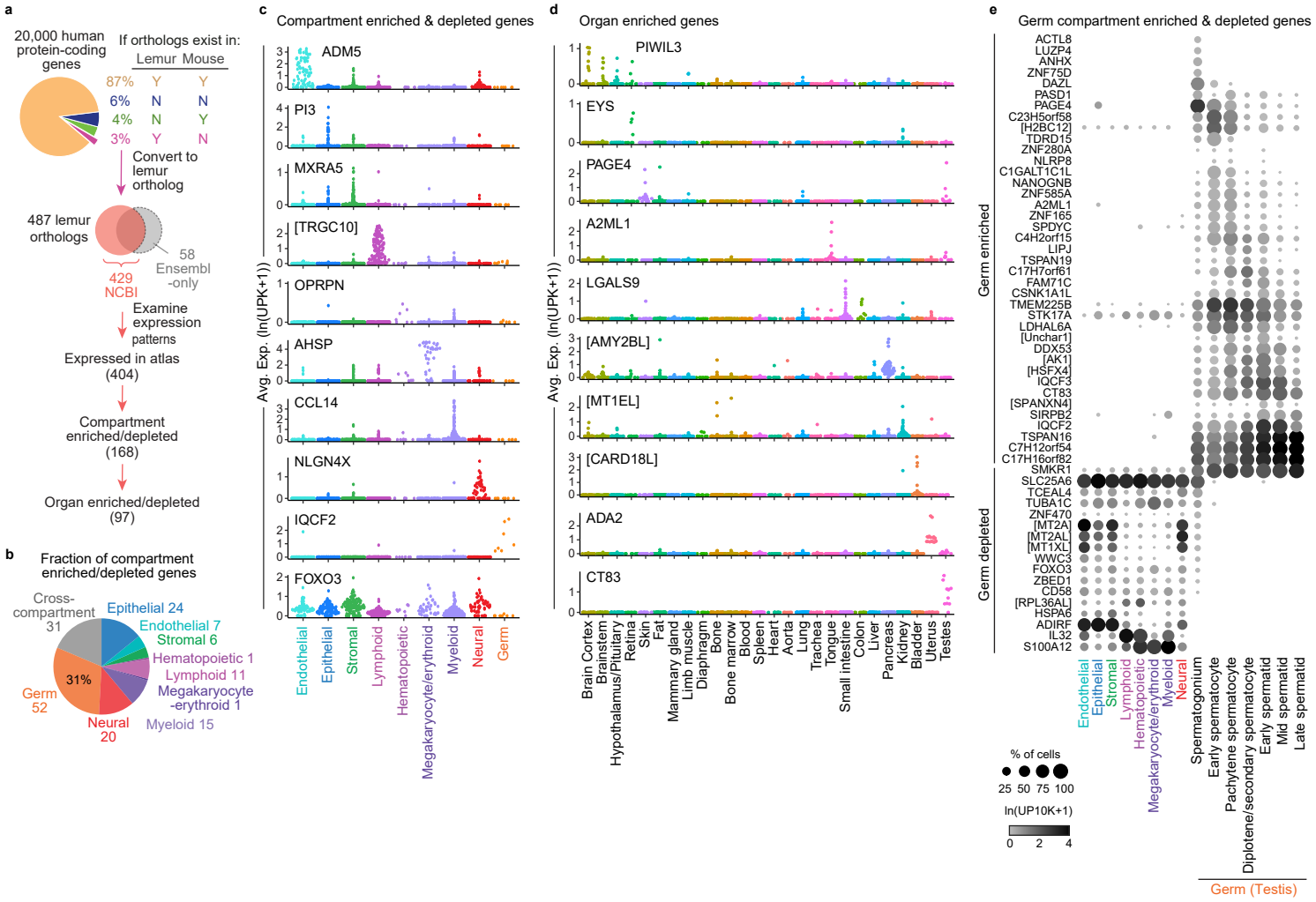

a
20,000 human protein-coding genes
If orthologs exist in:
Lemur Mouse
87% Y Y
6% N N
4% N Y
3% Y N
Convert to lemur ortholog
487 lemur orthologs
429 NCBI
58 Ensembl -only
Examine expression patterns
Expressed in atlas (404)
Compartment enriched/depleted (168)
Organ enriched/depleted (97)
b
Fraction of compartment enriched/depleted genes
Cross-compartment 31
Epithelial 24
Endothelial 7
Stromal 6
Hematopoietic 1
Lymphoid 11
Megakaryocyte -erythroid 1
Myeloid 15
Neural 20
Germ 52
31%
c
Compartment enriched & depleted genes
ADM5
PI3
MXRA5
[TRGC10]
OPRPN
AHSP
CCL14
NLGN4X
IQCF2
FOXO3
Avg. Exp. (ln(UPK+1))
Endothelial
Epithelial
Stromal
Lymphoid
Hematopoietic
Megakaryocyte/erythroid
Myeloid
Neural
Germ
d
Organ enriched genes
PIWIL3
EYS
PAGE4
A2ML1
LGALS9
[AMY2BL]
[MT1EL]
[CARD18L]
ADA2
CT83
Avg. Exp. (ln(UPK+1))
Brain Cortex
Brainstem
Hypothalamus/Pituitary
Retina
Skin
Fat
Mammary gland
Limb muscle
Diaphragm
Bone
Bone marrow
Blood
Spleen
Heart
Aorta
Lung
Trachea
Tongue
Small intestine
Colon
Liver
Pancreas
Kidney
Bladder
Uterus
Testes
e
Germ compartment enriched & depleted genes
Germ enriched
Germ depleted
ACTL8
LUZP4
ANHX
ZNF75D
DAZL
PASD1
PAGE4
C23H5orf58
[H2BC12]
TDRD15
ZNF280A
NLRP8
C1GALT1C1L
NANOGNB
ZNF585A
A2ML1
ZNF165
SPDYC
C4H2orf15
LIPJ
TSPAN19
C17H7orf61
FAM71C
CSNK1A1L
TMEM225B
STK17A
LDHAL6A
[Unchar1]
DDX53
[AK1]
[HSFX4]
IQCF3
CT83
[SPANXN4]
SIRPB2
IQCF2
TSPAN16
C7H12orf54
C17H16orf82
SMKR1
SLC25A6
TCEAL4
TUBA1C
ZNF470
[MT2A]
[MT2AL]
[MT1XL]
WWC3
FOXO3
ZBED1
CD58
[RPL36AL]
HSPA6
ADIRF
IL32
S100A12
Endothelial
Epithelial
Stromal
Lymphoid
Hematopoietic
Megakaryocyte/erythroid
Myeloid
Neural
Spermatogonium
Early spermatocyte
Pachytene spermatocyte
Diplotene/secondary spermatocyte
Early spermatid
Mid spermatid
Late spermatid
Germ (Testis)
% of cells
25 50 75 100
ln(UP10K+1)
0 2 4

Fig. S1
uTAR

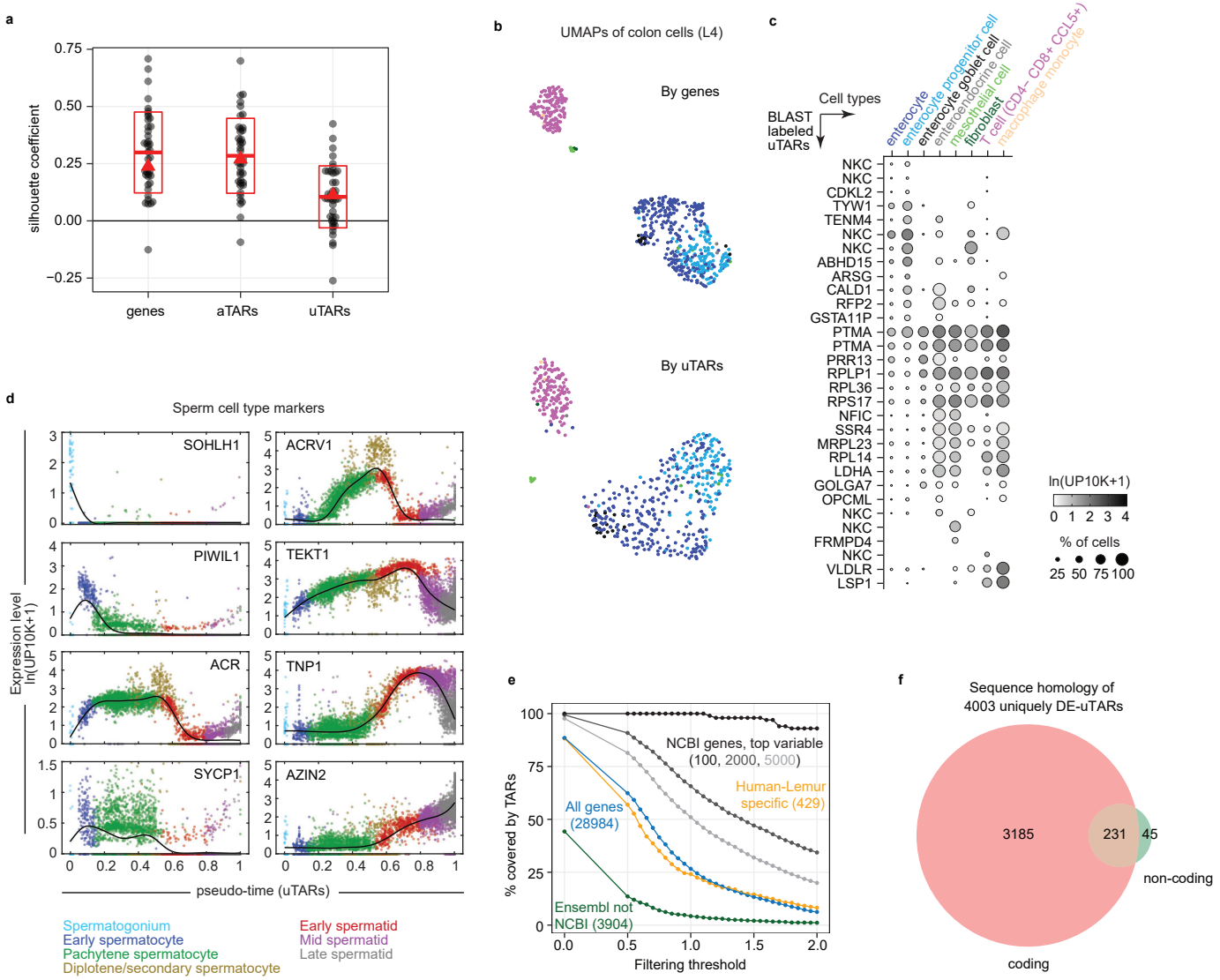

a
silhouette coefficient
genes
aTARs
uTARs
b
UMAPs of colon cells (L4)
By genes
By uTARs
c
Cell types
BLAST labeled uTARs
enterocyte
enterocyte progenitor cell
enterocyte goblet cell
enteroendocrine cell
mesothelial cell
fibroblast
T cell (CD4- CD8+ CCL5+)
macrophage monocyte
ln(UP10K+1)
% of cells
d
Sperm cell type markers
SOHLH1
ACRV1
PIWIL1
TEKT1
ACR
TNP1
SYCP1
AZIN2
Expression level ln(UP10K+1)
pseudo-time (uTARs)
Spermatogonium
Early spermatocyte
Pachytene spermatocyte
Diplotene/secondary spermatocyte
Early spermatid
Mid spermatid
Late spermatid
e
% covered by TARs
Filtering threshold
NCBI genes, top variable (100, 2000, 5000)
Human-Lemur specific (429)
All genes (28984)
Ensembl not NCBI (3904)
f
Sequence homology of 4003 uniquely DE-uTARs
3185
231
45
non-coding
coding

Fig. S2

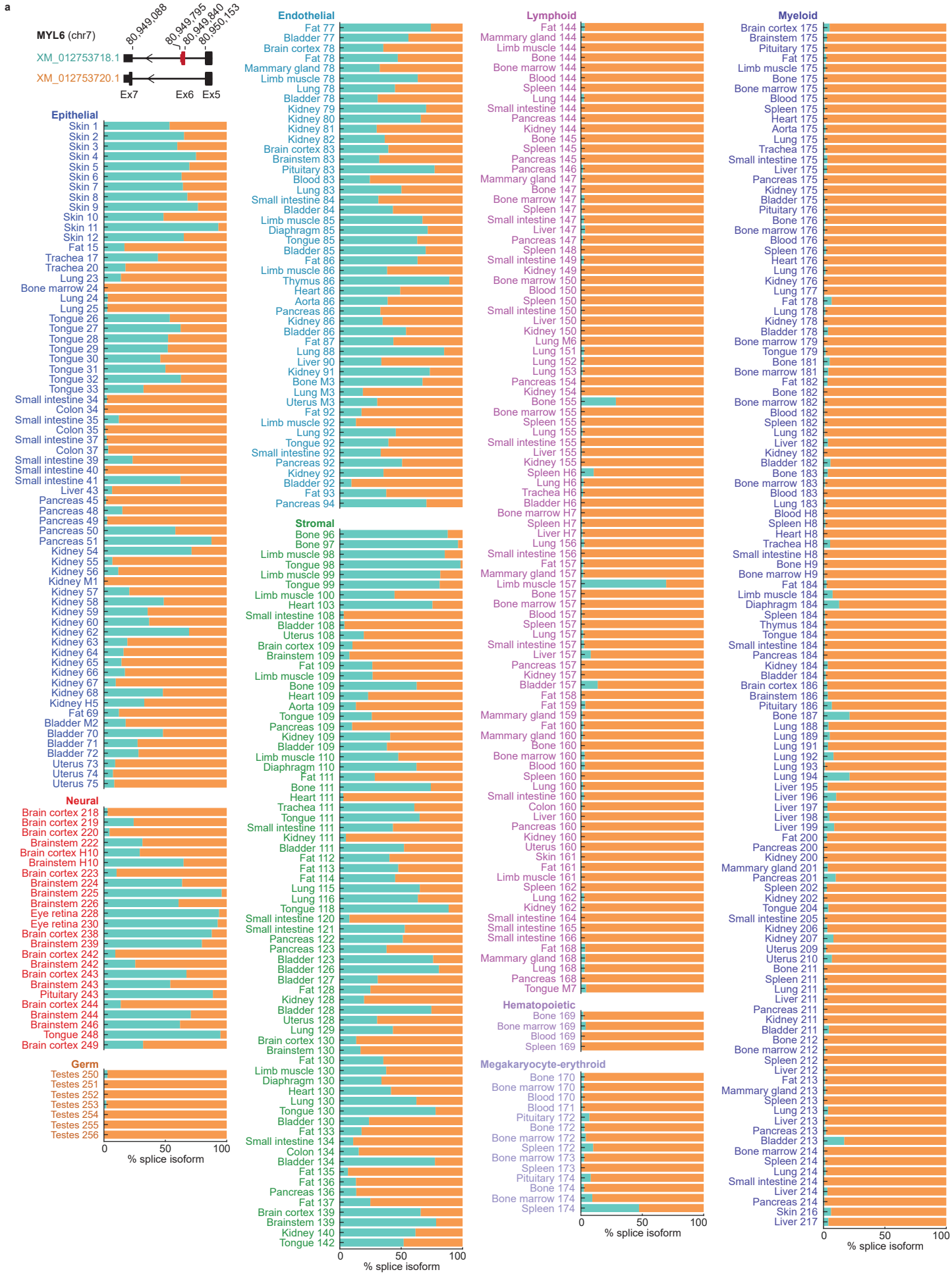
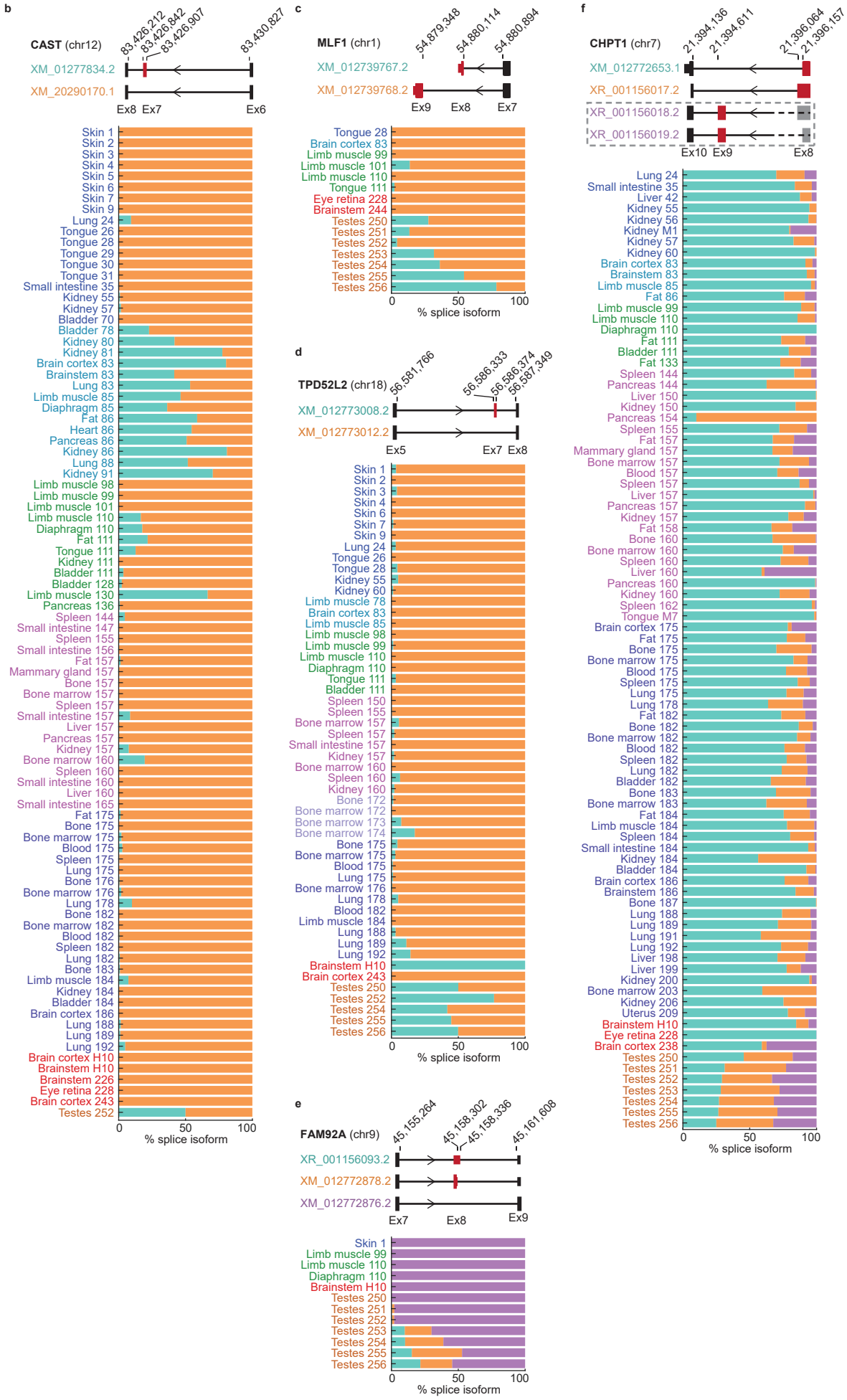
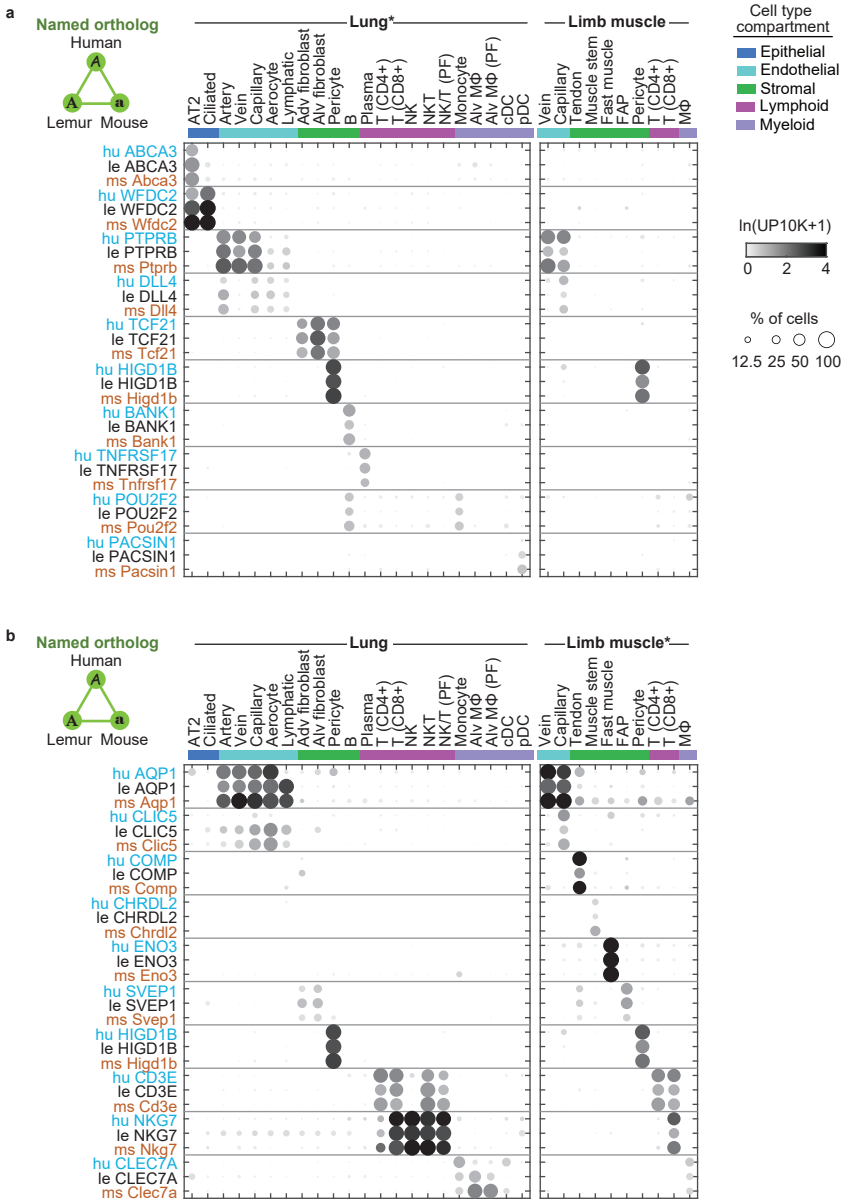

a
MYL6 (chr7)
80,949,088
80,949,795
80,949,840
80,950,153
XM_012753718.1
XM_012753720.1
Ex7
Ex6
Ex5
Epithelial
Skin 1
Skin 2
Skin 3
Skin 4
Skin 5
Skin 6
Skin 7
Skin 8
Skin 9
Skin 10
Skin 11
Skin 12
Fat 15
Trachea 17
Trachea 20
Lung 23
Bone marrow 24
Lung 24
Lung 25
Tongue 26
Tongue 27
Tongue 28
Tongue 29
Tongue 30
Tongue 31
Tongue 32
Tongue 33
Small intestine 34
Colon 34
Small intestine 35
Colon 35
Small intestine 37
Colon 37
Small intestine 39
Small intestine 40
Small intestine 41
Liver 43
Pancreas 45
Pancreas 48
Pancreas 49
Pancreas 50
Pancreas 51
Kidney 54
Kidney 55
Kidney 56
Kidney M1
Kidney 57
Kidney 58
Kidney 59
Kidney 60
Kidney 62
Kidney 63
Kidney 64
Kidney 65
Kidney 66
Kidney 67
Kidney 68
Kidney H5
Fat 69
Bladder M2
Bladder 70
Bladder 71
Bladder 72
Uterus 73
Uterus 74
Uterus 75
Neural
Brain cortex 218
Brain cortex 219
Brain cortex 220
Brainstem 222
Brain cortex H10
Brainstem H10
Brain cortex 223
Brainstem 224
Brainstem 225
Brainstem 226
Eye retina 228
Eye retina 230
Brain cortex 238
Brainstem 239
Brain cortex 242
Brainstem 242
Brain cortex 243
Brainstem 243
Pituitary 243
Brain cortex 244
Brainstem 244
Brainstem 246
Tongue 248
Brain cortex 249
Germ
Testes 250
Testes 251
Testes 252
Testes 253
Testes 254
Testes 255
Testes 256
0
50
100
% splice isoform
Endothelial
Fat 77
Bladder 77
Brain cortex 78
Fat 78
Mammary gland 78
Limb muscle 78
Lung 78
Bladder 78
Kidney 79
Kidney 80
Kidney 81
Kidney 82
Brain cortex 83
Brainstem 83
Pituitary 83
Blood 83
Lung 83
Small intestine 84
Bladder 84
Limb muscle 85
Diaphragm 85
Tongue 85
Bladder 85
Fat 86
Limb muscle 86
Thymus 86
Heart 86
Aorta 86
Pancreas 86
Kidney 86
Bladder 86
Fat 87
Lung 88
Liver 90
Kidney 91
Bone M3
Lung M3
Uterus M3
Fat 92
Limb muscle 92
Lung 92
Tongue 92
Small intestine 92
Pancreas 92
Kidney 92
Bladder 92
Fat 93
Pancreas 94
Stromal
Bone 96
Bone 97
Limb muscle 98
Tongue 98
Limb muscle 99
Tongue 99
Limb muscle 100
Heart 103
Small intestine 108
Bladder 108
Uterus 108
Brain cortex 109
Brainstem 109
Fat 109
Limb muscle 109
Bone 109
Heart 109
Aorta 109
Tongue 109
Pancreas 109
Kidney 109
Bladder 109
Limb muscle 110
Diaphragm 110
Fat 111
Bone 111
Heart 111
Trachea 111
Tongue 111
Small intestine 111
Kidney 111
Bladder 111
Fat 112
Fat 113
Fat 114
Lung 115
Lung 116
Tongue 118
Small intestine 120
Small intestine 121
Pancreas 122
Pancreas 123
Bladder 123
Bladder 126
Bladder 127
Fat 128
Kidney 128
Bladder 128
Uterus 128
Lung 129
Brain cortex 130
Brainstem 130
Fat 130
Limb muscle 130
Diaphragm 130
Heart 130
Lung 130
Tongue 130
Bladder 130
Fat 133
Small intestine 134
Colon 134
Bladder 134
Fat 135
Fat 136
Pancreas 136
Fat 137
Brain cortex 139
Brainstem 139
Kidney 140
Tongue 142
0
50
100
% splice isoform
Lymphoid
Fat 144
Mammary gland 144
Limb muscle 144
Bone 144
Bone marrow 144
Blood 144
Spleen 144
Lung 144
Small intestine 144
Pancreas 144
Kidney 144
Bone 145
Spleen 145
Pancreas 145
Pancreas 146
Mammary gland 147
Bone 147
Bone marrow 147
Spleen 147
Small intestine 147
Liver 147
Pancreas 147
Spleen 148
Small intestine 149
Kidney 149
Bone marrow 150
Blood 150
Spleen 150
Small intestine 150
Liver 150
Kidney 150
Lung M6
Lung 151
Lung 152
Lung 153
Pancreas 154
Kidney 154
Bone 155
Bone marrow 155
Spleen 155
Lung 155
Small intestine 155
Liver 155
Kidney 155
Spleen H6
Lung H6
Trachea H6
Bladder H6
Bone marrow H7
Spleen H7
Liver H7
Lung 156
Small intestine 156
Fat 157
Mammary gland 157
Limb muscle 157
Bone 157
Bone marrow 157
Blood 157
Spleen 157
Lung 157
Small intestine 157
Liver 157
Pancreas 157
Kidney 157
Bladder 157
Fat 158
Fat 159
Mammary gland 159
Fat 160
Mammary gland 160
Bone 160
Bone marrow 160
Blood 160
Spleen 160
Lung 160
Small intestine 160
Colon 160
Liver 160
Pancreas 160
Kidney 160
Uterus 160
Skin 161
Fat 161
Limb muscle 161
Spleen 162
Lung 162
Kidney 162
Small intestine 164
Small intestine 165
Small intestine 166
Fat 168
Mammary gland 168
Lung 168
Pancreas 168
Tongue M7
Hematopoietic
Bone 169
Bone marrow 169
Blood 169
Spleen 169
Megakaryocyte-erythroid
Bone 170
Bone marrow 170
Blood 170
Blood 171
Pituitary 172
Bone 172
Bone marrow 172
Spleen 172
Bone marrow 173
Spleen 173
Pituitary 174
Bone 174
Bone marrow 174
Spleen 174
0
50
100
% splice isoform
Myeloid
Brain cortex 175
Brainstem 175
Pituitary 175
Fat 175
Limb muscle 175
Bone 175
Bone marrow 175
Blood 175
Spleen 175
Heart 175
Aorta 175
Lung 175
Trachea 175
Small intestine 175
Liver 175
Pancreas 175
Kidney 175
Bladder 175
Pituitary 176
Bone 176
Bone marrow 176
Blood 176
Spleen 176
Heart 176
Lung 176
Kidney 176
Lung 177
Fat 178
Lung 178
Kidney 178
Bladder 178
Bone marrow 179
Tongue 179
Bone 181
Bone marrow 181
Fat 182
Bone 182
Bone marrow 182
Blood 182
Spleen 182
Lung 182
Liver 182
Kidney 182
Bladder 182
Bone 183
Bone marrow 183
Blood 183
Lung 183
Blood H8
Spleen H8
Heart H8
Trachea H8
Small intestine H8
Bone H9
Bone marrow H9
Fat 184
Limb muscle 184
Diaphragm 184
Spleen 184
Thymus 184
Tongue 184
Small intestine 184
Pancreas 184
Kidney 184
Bladder 184
Brain cortex 186
Brainstem 186
Pituitary 186
Bone 187
Lung 188
Lung 189
Lung 191
Lung 192
Lung 193
Lung 194
Liver 195
Liver 196
Liver 197
Liver 198
Liver 199
Fat 200
Pancreas 200
Kidney 200
Mammary gland 201
Pancreas 201
Spleen 202
Kidney 202
Tongue 204
Small intestine 205
Kidney 206
Kidney 207
Uterus 209
Uterus 210
Bone 211
Spleen 211
Lung 211
Liver 211
Pancreas 211
Kidney 211
Bladder 211
Bone 212
Bone marrow 212
Spleen 212
Liver 212
Fat 213
Mammary gland 213
Spleen 213
Lung 213
Liver 213
Pancreas 213
Bladder 213
Bone marrow 214
Spleen 214
Lung 214
Small intestine 214
Liver 214
Pancreas 214
Skin 216
Liver 217
0
50
100
% splice isoform

Fig. S2 - continued




a
Named ortholog
Human
Lemur
Mouse
Lung*
Limb muscle
AT2
Ciliated
Artery
Vein
Capillary
Aerocyte
Lymphatic
Adv fibroblast
Alv fibroblast
Pericyte
B
Plasma
T (CD4+)
T (CD8+)
NK
NKT
NK/T (PF)
Monocyte
Alv MΦ
Alv MΦ (PF)
cDC
pDC
Vein
Capillary
Tendon
Muscle stem
Fast muscle
FAP
Pericyte
T (CD4+)
T (CD8+)
MΦ
hu ABCA3
le ABCA3
ms Abca3
hu WFDC2
le WFDC2
ms Wfdc2
hu PTPRB
le PTPRB
ms Ptprb
hu DLL4
le DLL4
ms Dll4
hu TCF21
le TCF21
ms Tcf21
hu HIGD1B
le HIGD1B
ms Higd1b
hu BANK1
le BANK1
ms Bank1
hu TNFRSF17
le TNFRSF17
ms Tnfrsf17
hu POU2F2
le POU2F2
ms Pou2f2
hu PACSIN1
le PACSIN1
ms Pacsin1
Cell type compartment
Epithelial
Endothelial
Stromal
Lymphoid
Myeloid
ln(UP10K+1)
0
2
4
% of cells
12.5 25 50 100

b
Named ortholog
Human
Lemur
Mouse
Lung
Limb muscle*
AT2
Ciliated
Artery
Vein
Capillary
Aerocyte
Lymphatic
Adv fibroblast
Alv fibroblast
Pericyte
B
Plasma
T (CD4+)
T (CD8+)
NK
NKT
NK/T (PF)
Monocyte
Alv MΦ
Alv MΦ (PF)
cDC
pDC
Vein
Capillary
Tendon
Muscle stem
Fast muscle
FAP
Pericyte
T (CD4+)
T (CD8+)
MΦ
hu AQP1
le AQP1
ms Aqp1
hu CLIC5
le CLIC5
ms Clic5
hu COMP
le COMP
ms Comp
hu CHRDL2
le CHRDL2
ms Chrdl2
hu ENO3
le ENO3
ms Eno3
hu SVEP1
le SVEP1
ms Svep1
hu HIGD1B
le HIGD1B
ms Higd1b
hu CD3E
le CD3E
ms Cd3e
hu NKG7
le NKG7
ms Nkg7
hu CLEC7A
le CLEC7A
ms Clec7a

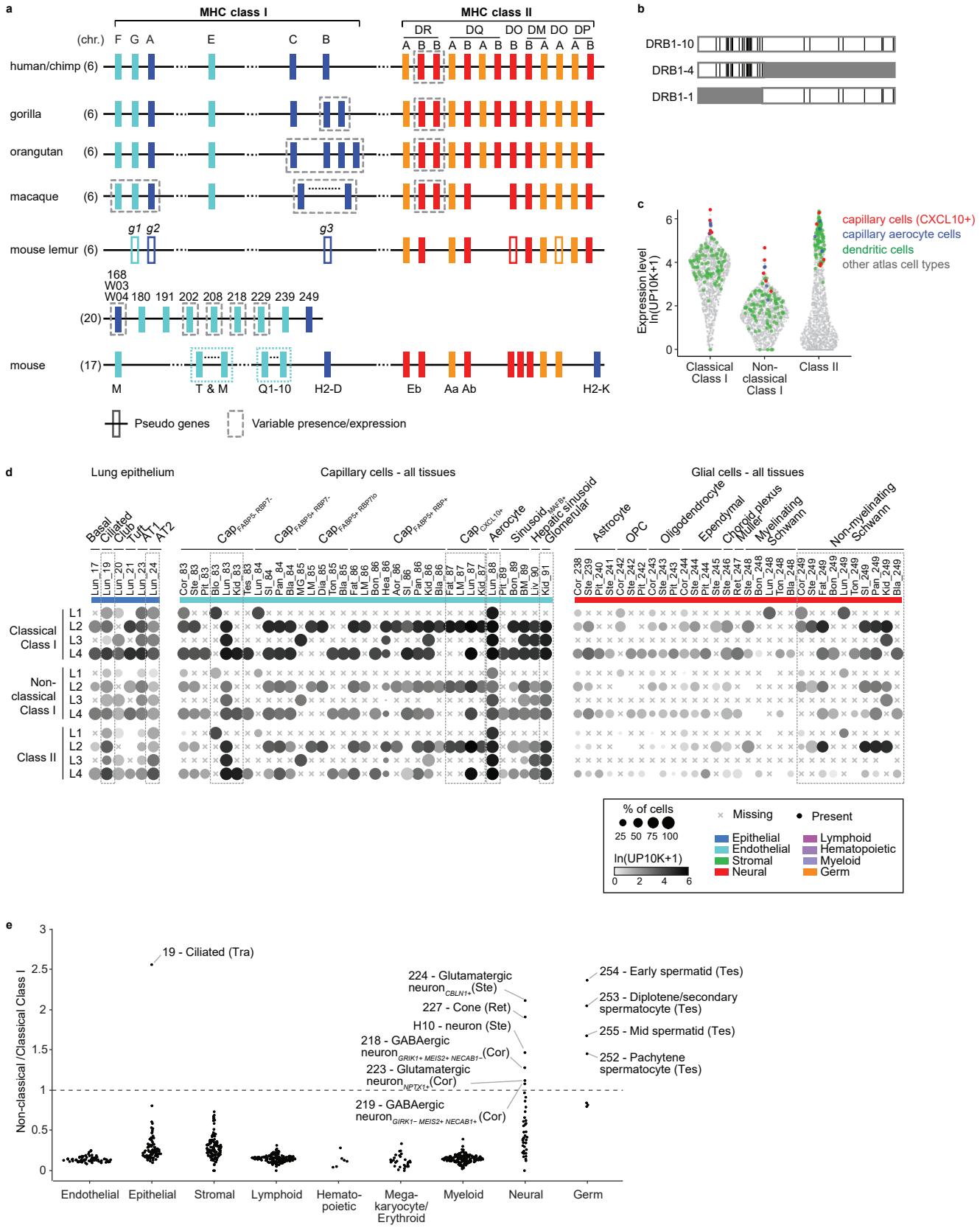

a
MHC class I
MHC class II
(chr.) F G A E C B
DR DQ DO DM DO DP
A B B A B A B B B A A A B
human/chimp (6)
gorilla (6)
orangutan (6)
macaque (6)
mouse lemur (6)
g1 g2 g3
168 W03 W04 180 191 202 208 218 229 239 249
(20)
mouse (17)
M T & M Q1-10 H2-D Eb Aa Ab H2-K
Pseudo genes
Variable presence/expression
b
DRB1-10
DRB1-4
DRB1-1
c
Expression level ln(UP10K+1)
capillary cells (CXCL10+)
capillary aerocyte cells
dendritic cells
other atlas cell types
Classical Class I
Non-classical Class I
Class II
d
Lung epithelium
Capillary cells - all tissues
Glial cells - all tissues
Basal
Ciliated
Club
Tuft
AT1
AT2
Aerocyte
Sinusoid
Hepatic sinusoid
Glomerular
Astrocyte
OPC
Oligodendrocyte
Ependymal
Choroid plexus
Muller
Myelinating Schwann
Non-myelinating Schwann
Classical Class I
Non-classical Class I
Class II
L1
L2
L3
L4
% of cells
25 50 75 100
ln(UP10K+1)
0 2 4 6
Missing
Present
Epithelial
Endothelial
Stromal
Neural
Lymphoid
Hematopoietic
Myeloid
Germ
e
Non-classical /Classical Class I
19 - Ciliated (Tra)
224 - Glutamatergic neuron CBLN1+ (Ste)
227 - Cone (Ret)
H10 - neuron (Ste)
218 - GABAergic neuron GRIK1+ MEIS2+ NECAB1− (Cor)
223 - Glutamatergic neuron NPTX1+ (Cor)
219 - GABAergic neuron GIRK1− MEIS2+ NECAB1+ (Cor)
254 - Early spermatid (Tes)
253 - Diplotene/secondary spermatocyte (Tes)
255 - Mid spermatid (Tes)
252 - Pachytene spermatocyte (Tes)
Endothelial
Epithelial
Stromal
Lymphoid
Hemato-poietic
Mega-karyocyte/ Erythroid
Myeloid
Neural
Germ

Fig. S5

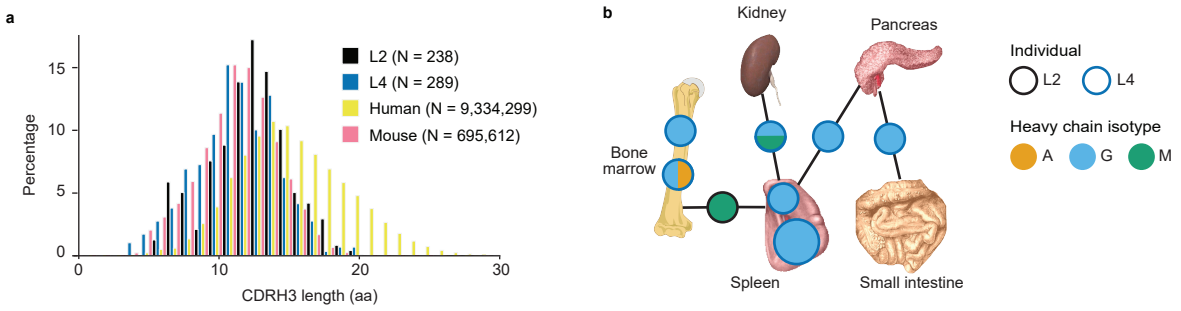

a
L2 (N = 238)
L4 (N = 289)
Human (N = 9,334,299)
Mouse (N = 695,612)
Percentage
0
5
10
15
0
10
20
30
CDRH3 length (aa)
b
Kidney
Pancreas
Bone marrow
Spleen
Small intestine
Individual
L2
L4
Heavy chain isotype
A
G
M

Fig S6

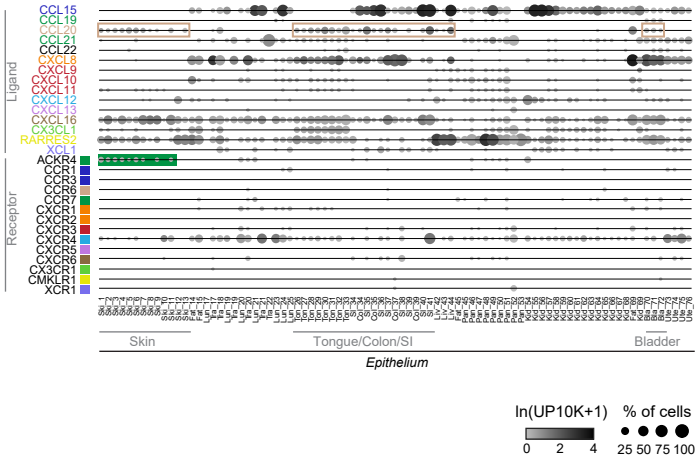

Ligand
CCL15
CCL19
CCL20
CCL21
CCL22
CXCL8
CXCL9
CXCL10
CXCL11
CXCL12
CXCL13
CXCL16
CX3CL1
RARRES2
XCL1
Receptor
ACKR4
CCR1
CCR3
CCR6
CCR7
CXCR1
CXCR2
CXCR3
CXCR4
CXCR5
CXCR6
CX3CR1
CMKLR1
XCR1
Skin
Tongue/Colon/SI
Bladder
Epithelium
ln(UP10K+1)
0 2 4
% of cells
25 50 75 100

Fig S7

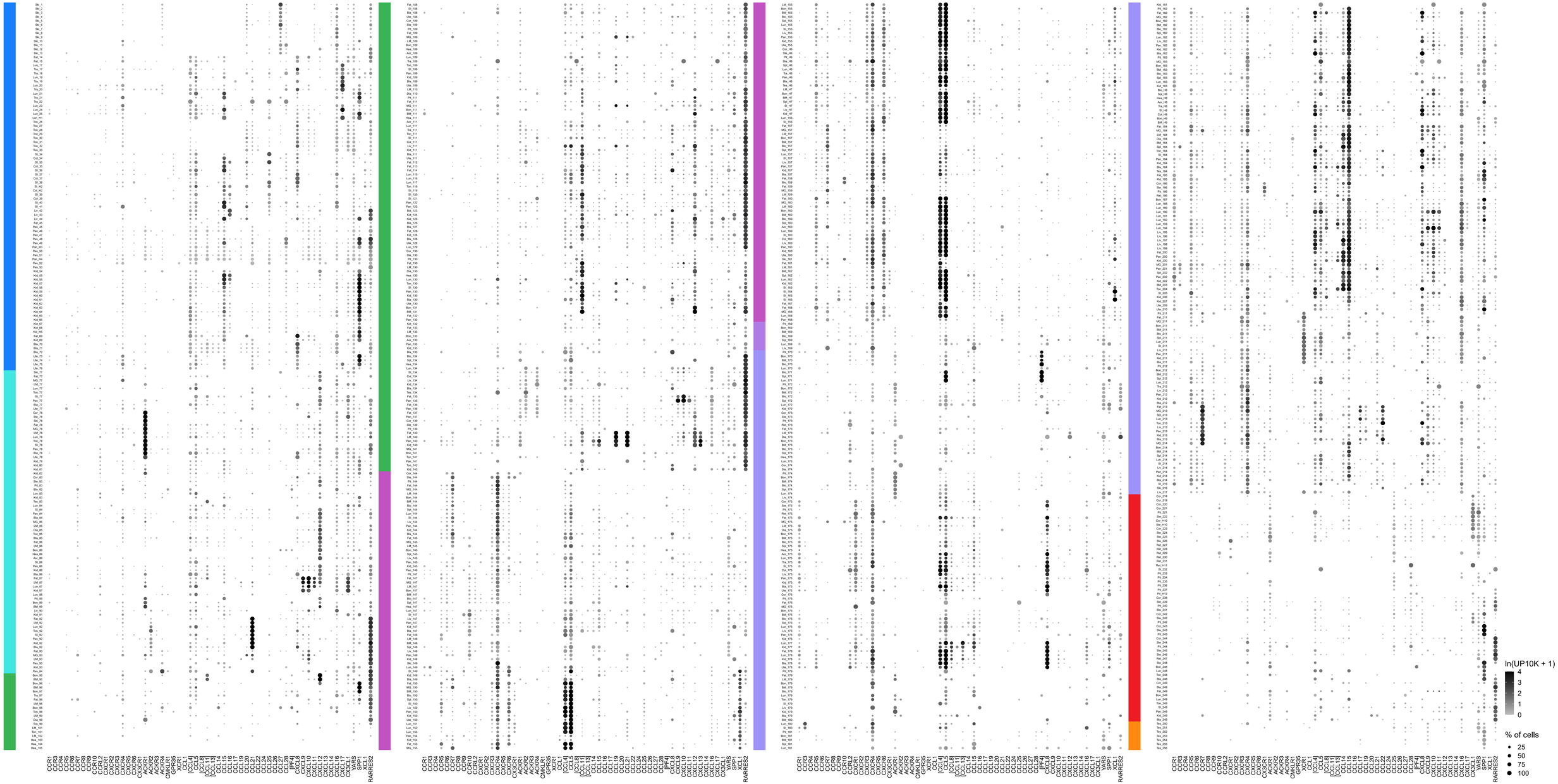
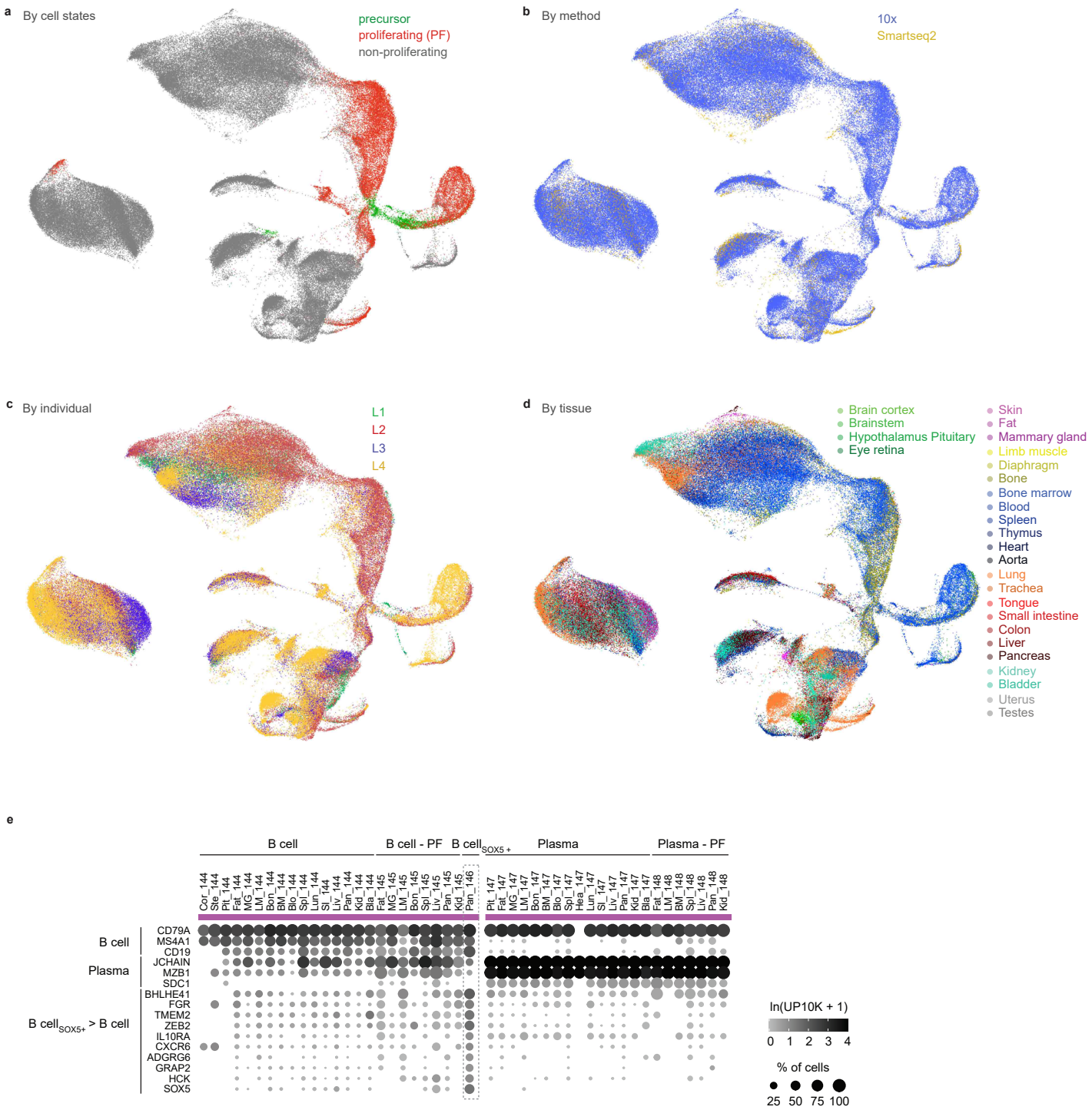
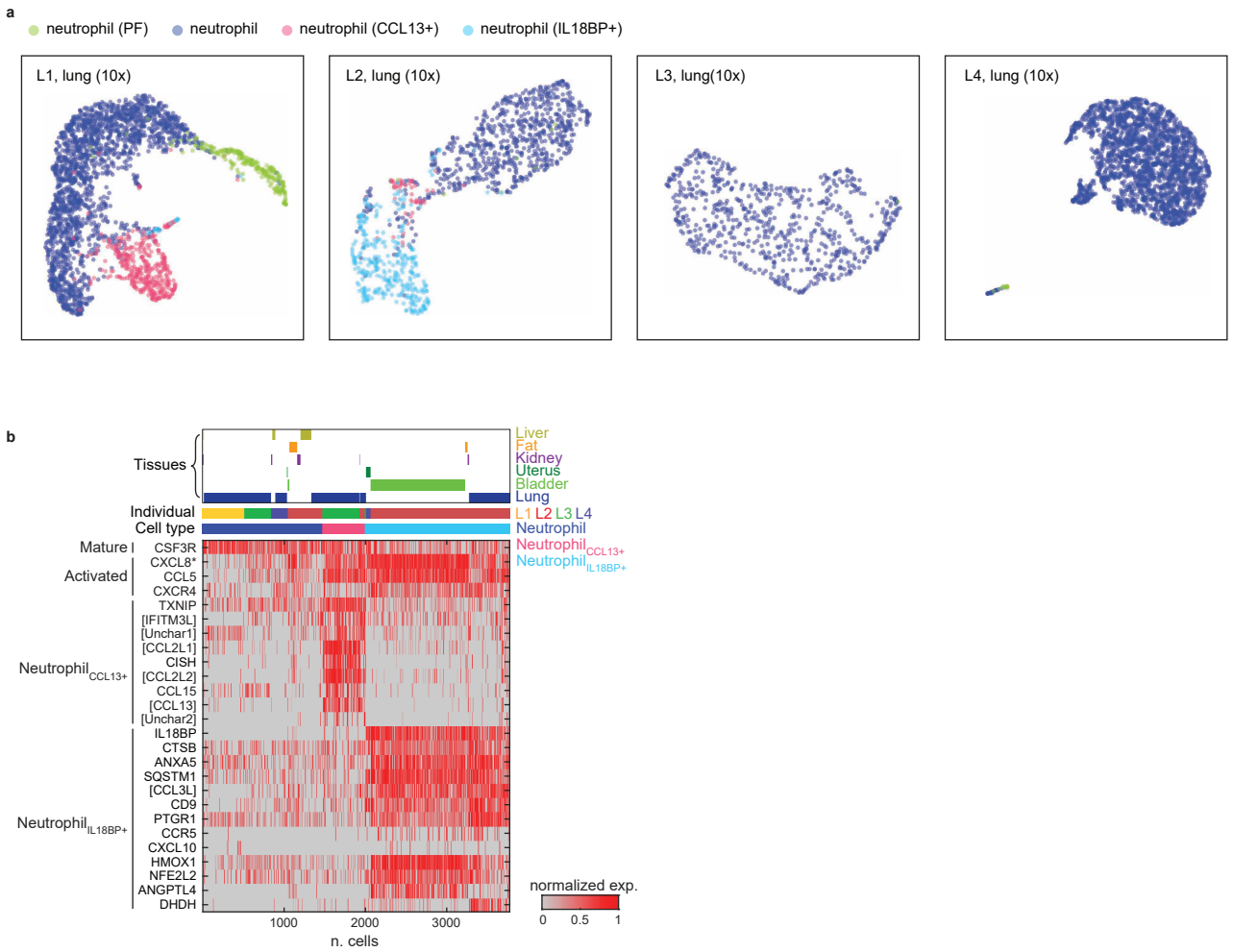
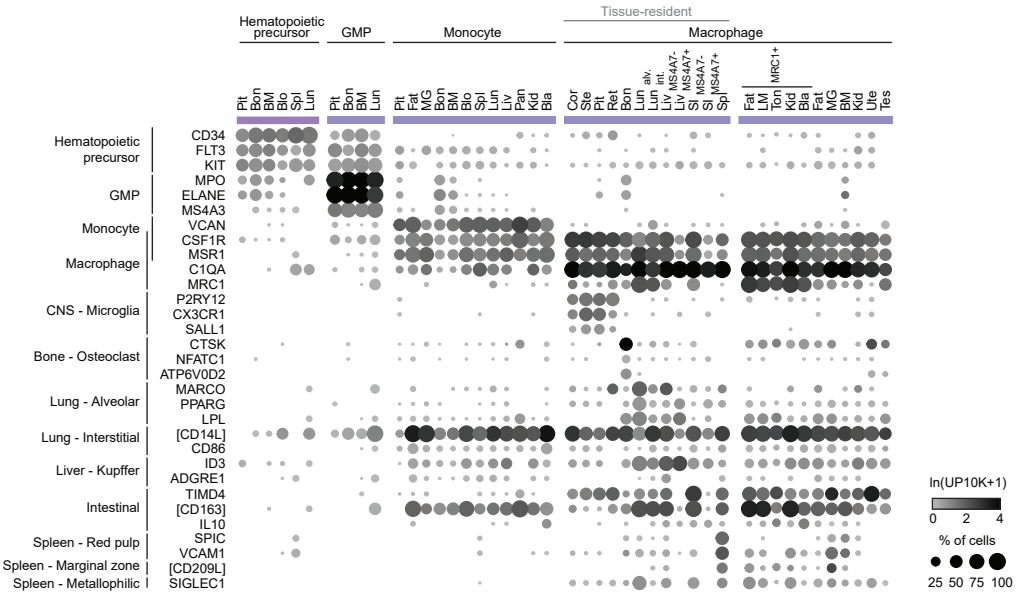


Fig. S8

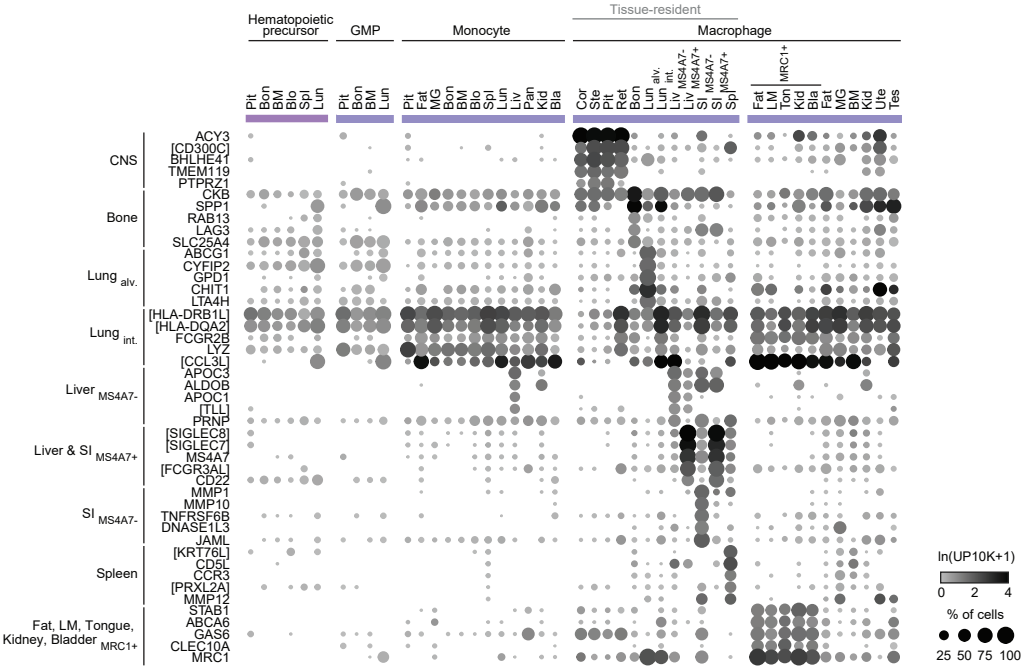

a
By cell states
precursor
proliferating (PF)
non-proliferating
b
By method
10x
Smartseq2
c
By individual
L1
L2
L3
L4
d
By tissue
Brain cortex
Brainstem
Hypothalamus Pituitary
Eye retina
Skin
Fat
Mammary gland
Limb muscle
Diaphragm
Bone
Bone marrow
Blood
Spleen
Thymus
Heart
Aorta
Lung
Trachea
Tongue
Small intestine
Colon
Liver
Pancreas
Kidney
Bladder
Uterus
Testes
e
B cell
B cell - PF
B cell SOX5 +
Plasma
Plasma - PF
Cor_144
Ste_144
Pit_144
Fat_144
MG_144
LM_144
Bon_144
BM_144
Blo_144
Spl_144
Lun_144
Sl_144
Liv_144
Pan_144
Kid_144
Bla_144
Fat_145
MG_145
LM_145
Bon_145
Spl_145
Liv_145
Pan_145
Kid_145
Pan_146
Pit_147
Fat_147
MG_147
LM_147
Bon_147
BM_147
Blo_147
Spl_147
Hea_147
Lun_147
Sl_147
Liv_147
Pan_147
Kid_147
Bla_147
Fat_148
LM_148
BM_148
Spl_148
Liv_148
Pan_148
Kid_148
B cell
Plasma
B cell SOX5+ > B cell
CD79A
MS4A1
CD19
JCHAIN
MZB1
SDC1
BHLHE41
FGR
TMEM2
ZEB2
IL10RA
CXCR6
ADGRG6
GRAP2
HCK
SOX5
ln(UP10K + 1)
0 1 2 3 4
% of cells
25 50 75 100


a
neutrophil (PF)
neutrophil
neutrophil (CCL13+)
neutrophil (IL18BP+)
L1, lung (10x)
L2, lung (10x)
L3, lung(10x)
L4, lung (10x)
b
Tissues
Liver
Fat
Kidney
Uterus
Bladder
Lung
Individual
L1 L2 L3 L4
Cell type
Neutrophil
Neutrophil CCL13+
Neutrophil IL18BP+
Mature
Activated
Neutrophil CCL13+
Neutrophil IL18BP+
CSF3R
CXCL8*
CCL5
CXCR4
TXNIP
[IFITM3L]
[Unchar1]
[CCL2L1]
CISH
[CCL2L2]
CCL15
[CCL13]
[Unchar2]
IL18BP
CTSB
ANXA5
SQSTM1
[CCL3L]
CD9
PTGR1
CCR5
CXCL10
HMOX1
NFE2L2
ANGPTL4
DHDH
1000
2000
3000
n. cells
normalized exp.
0
0.5
1

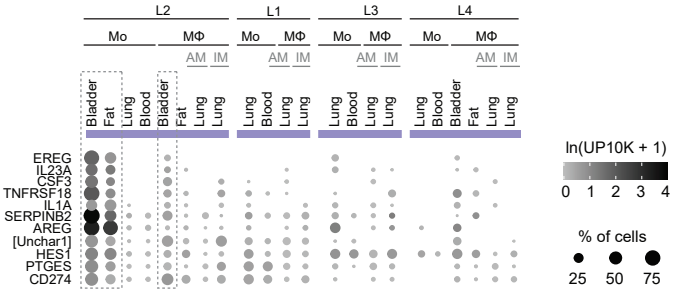
a Tissue-resident macrophage markers



b Tissue-resident macrophage differentially expressed genes



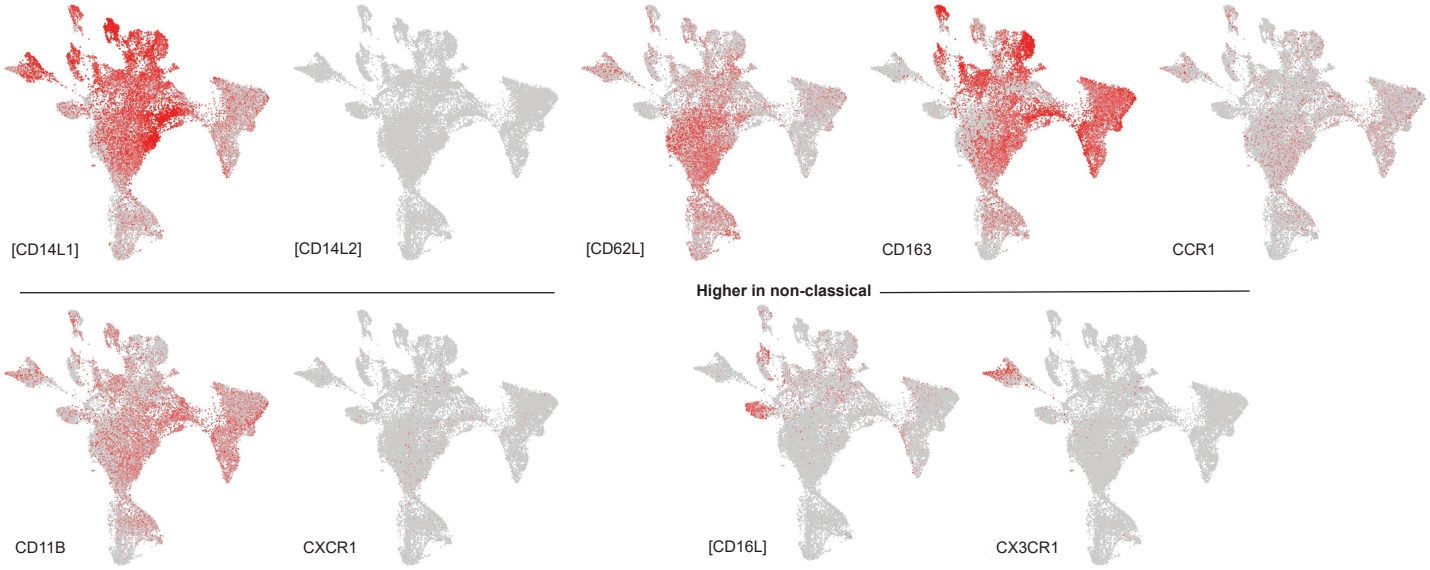
c Differentially expressed genes in activated monocytes/macrophages



d

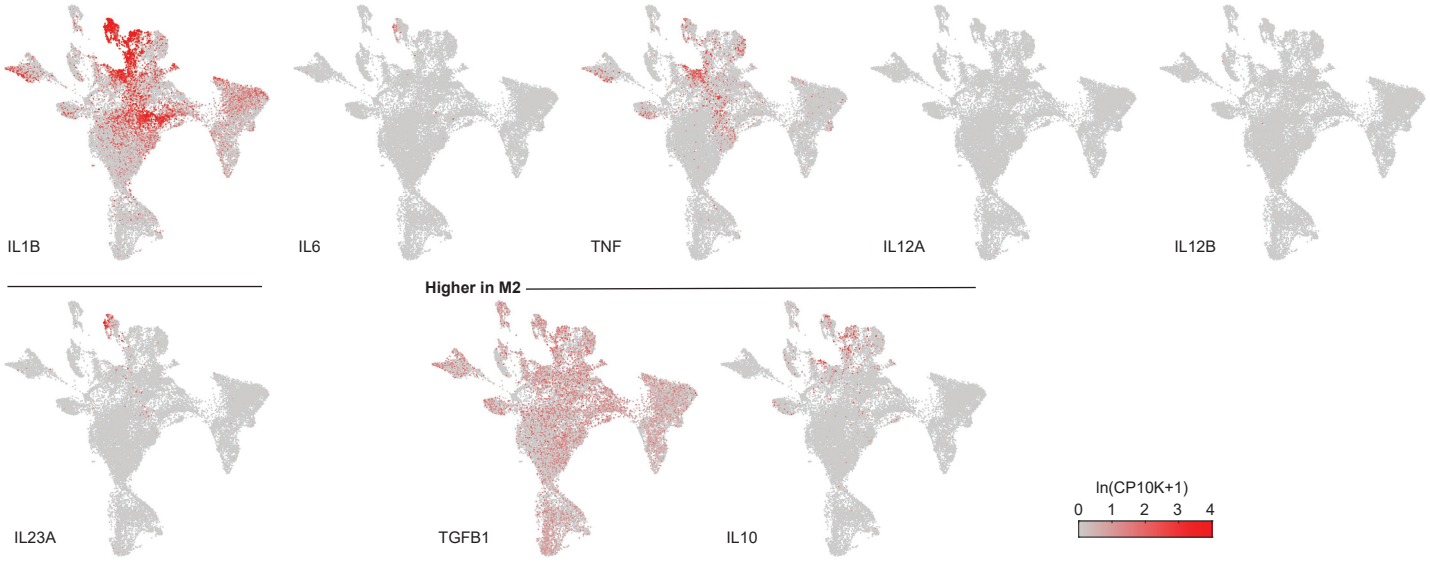
Monocyte classical vs non-classical markers

Higher in classical

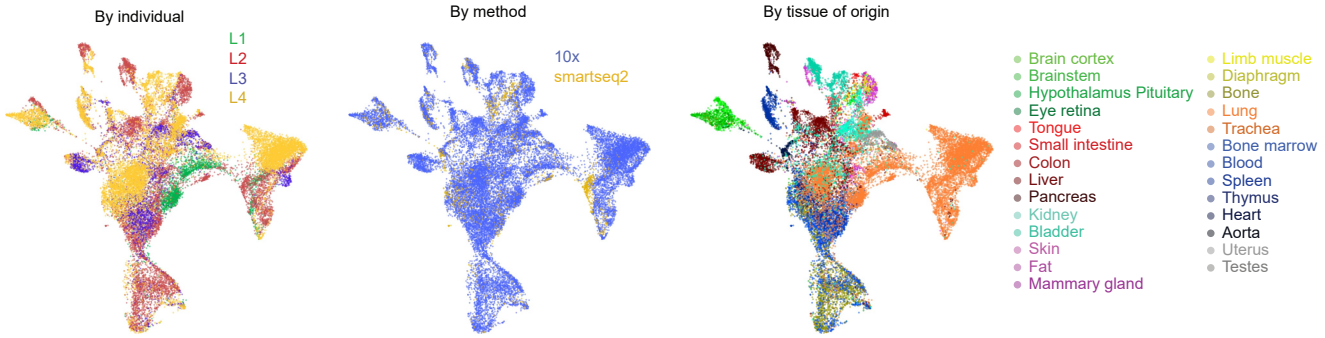


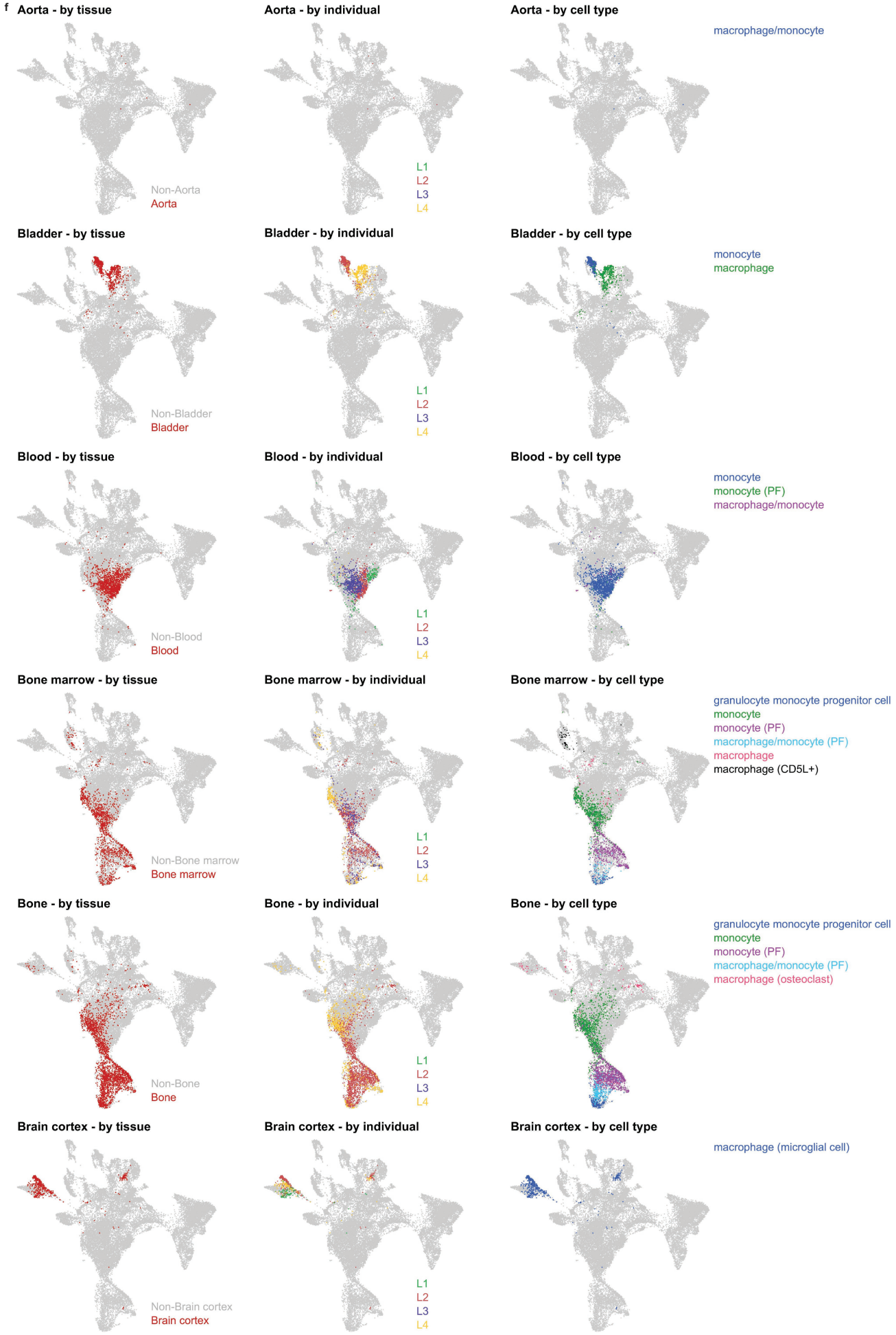
Macrophage M1/M2 markers

Higher in M1



e




f
Aorta - by tissue
Non-Aorta
Aorta
Aorta - by individual
L1
L2
L3
L4
Aorta - by cell type
macrophage/monocyte
Bladder - by tissue
Non-Bladder
Bladder
Bladder - by individual
L1
L2
L3
L4
Bladder - by cell type
monocyte
macrophage
Blood - by tissue
Non-Blood
Blood
Blood - by individual
L1
L2
L3
L4
Blood - by cell type
monocyte
monocyte (PF)
macrophage/monocyte
Bone marrow - by tissue
Non-Bone marrow
Bone marrow
Bone marrow - by individual
L1
L2
L3
L4
Bone marrow - by cell type
granulocyte monocyte progenitor cell
monocyte
monocyte (PF)
macrophage/monocyte (PF)
macrophage
macrophage (CD5L+)
Bone - by tissue
Non-Bone
Bone
Bone - by individual
L1
L2
L3
L4
Bone - by cell type
granulocyte monocyte progenitor cell
monocyte
monocyte (PF)
macrophage/monocyte (PF)
macrophage (osteoclast)
Brain cortex - by tissue
Non-Brain cortex
Brain cortex
Brain cortex - by individual
L1
L2
L3
L4
Brain cortex - by cell type
macrophage (microglial cell)

Brainstem - by tissue
Non-Brainstem
Brainstem
Brainstem - by individual
L1
L2
L3
L4
Brainstem - by cell type
macrophage (microglial cell)
Colon - by tissue
Non-Colon
Colon
Colon - by individual
L1
L2
L3
L4
Colon - by cell type
macrophage/monocyte
Diaphragm - by tissue
Non-Diaphragm
Diaphragm
Diaphragm - by individual
L1
L2
L3
L4
Diaphragm - by cell type
macrophage
Eye retina - by tissue
Non-Eye retina
Eye retina
Eye retina - by individual
L1
L2
L3
L4
Eye retina - by cell type
macrophage (microglial cell)
Fat - by tissue
Non-Fat
Fat
Fat - by individual
L1
L2
L3
L4
Fat - by cell type
monocyte
macrophage
macrophage (PF)
macrophage (MRC1+)
Heart - by tissue
Non-Heart
Heart
Heart - by individual
L1
L2
L3
L4
Heart - by cell type
macrophage/monocyte

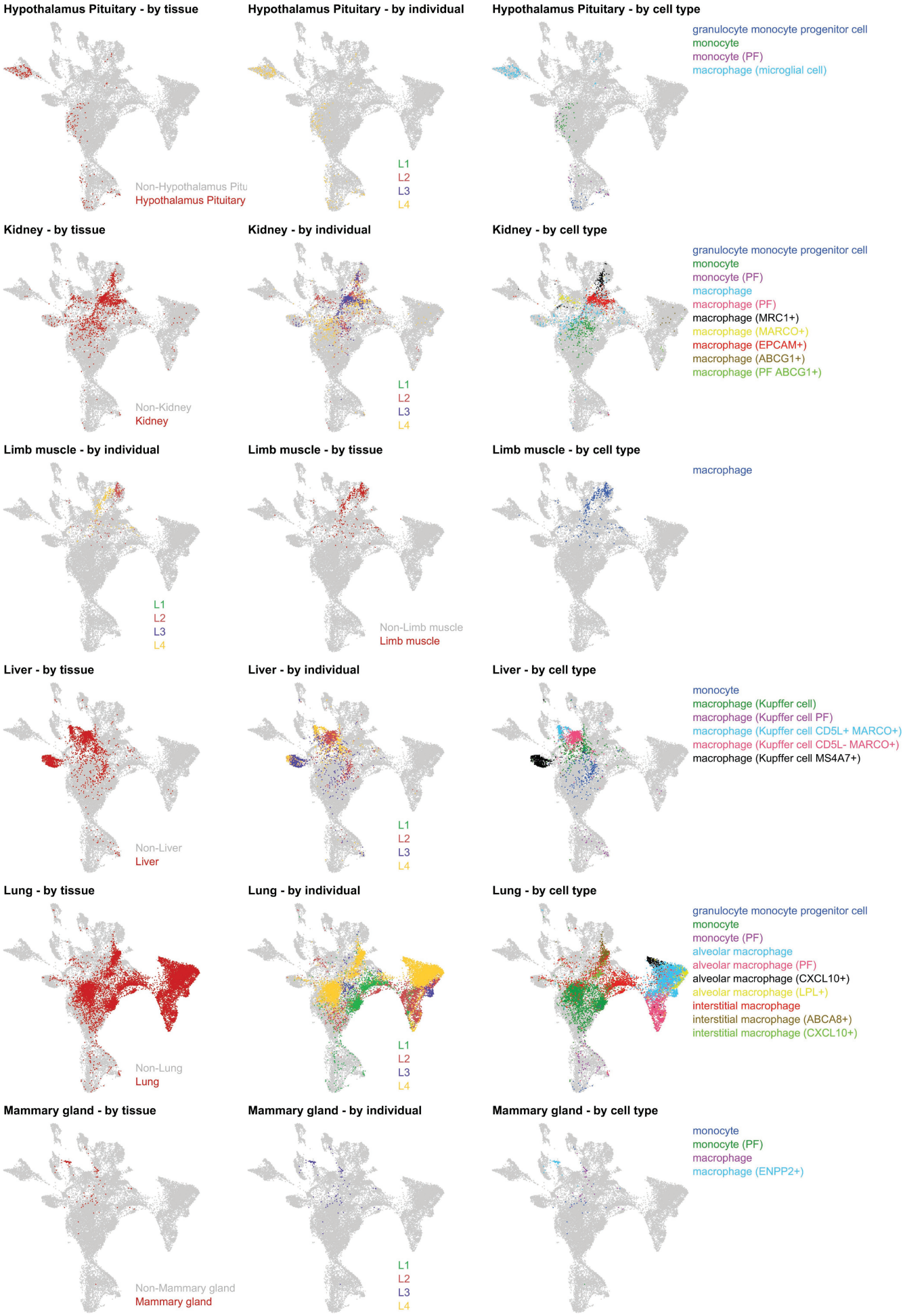

Hypothalamus Pituitary - by tissue
Non-Hypothalamus Pitu
Hypothalamus Pituitary
Hypothalamus Pituitary - by individual
L1
L2
L3
L4
Hypothalamus Pituitary - by cell type
granulocyte monocyte progenitor cell
monocyte
monocyte (PF)
macrophage (microglial cell)
Kidney - by tissue
Non-Kidney
Kidney
Kidney - by individual
L1
L2
L3
L4
Kidney - by cell type
granulocyte monocyte progenitor cell
monocyte
monocyte (PF)
macrophage
macrophage (PF)
macrophage (MRC1+)
macrophage (MARCO+)
macrophage (EPCAM+)
macrophage (ABCG1+)
macrophage (PF ABCG1+)
Limb muscle - by individual
L1
L2
L3
L4
Limb muscle - by tissue
Non-Limb muscle
Limb muscle
Limb muscle - by cell type
macrophage
Liver - by tissue
Non-Liver
Liver
Liver - by individual
L1
L2
L3
L4
Liver - by cell type
monocyte
macrophage (Kupffer cell)
macrophage (Kupffer cell PF)
macrophage (Kupffer cell CD5L+ MARCO+)
macrophage (Kupffer cell CD5L- MARCO+)
macrophage (Kupffer cell MS4A7+)
Lung - by tissue
Non-Lung
Lung
Lung - by individual
L1
L2
L3
L4
Lung - by cell type
granulocyte monocyte progenitor cell
monocyte
monocyte (PF)
alveolar macrophage
alveolar macrophage (PF)
alveolar macrophage (CXCL10+)
alveolar macrophage (LPL+)
interstitial macrophage
interstitial macrophage (ABCA8+)
interstitial macrophage (CXCL10+)
Mammary gland - by tissue
Non-Mammary gland
Mammary gland
Mammary gland - by individual
L1
L2
L3
L4
Mammary gland - by cell type
monocyte
monocyte (PF)
macrophage
macrophage (ENPP2+)

Fig. S10 - continued

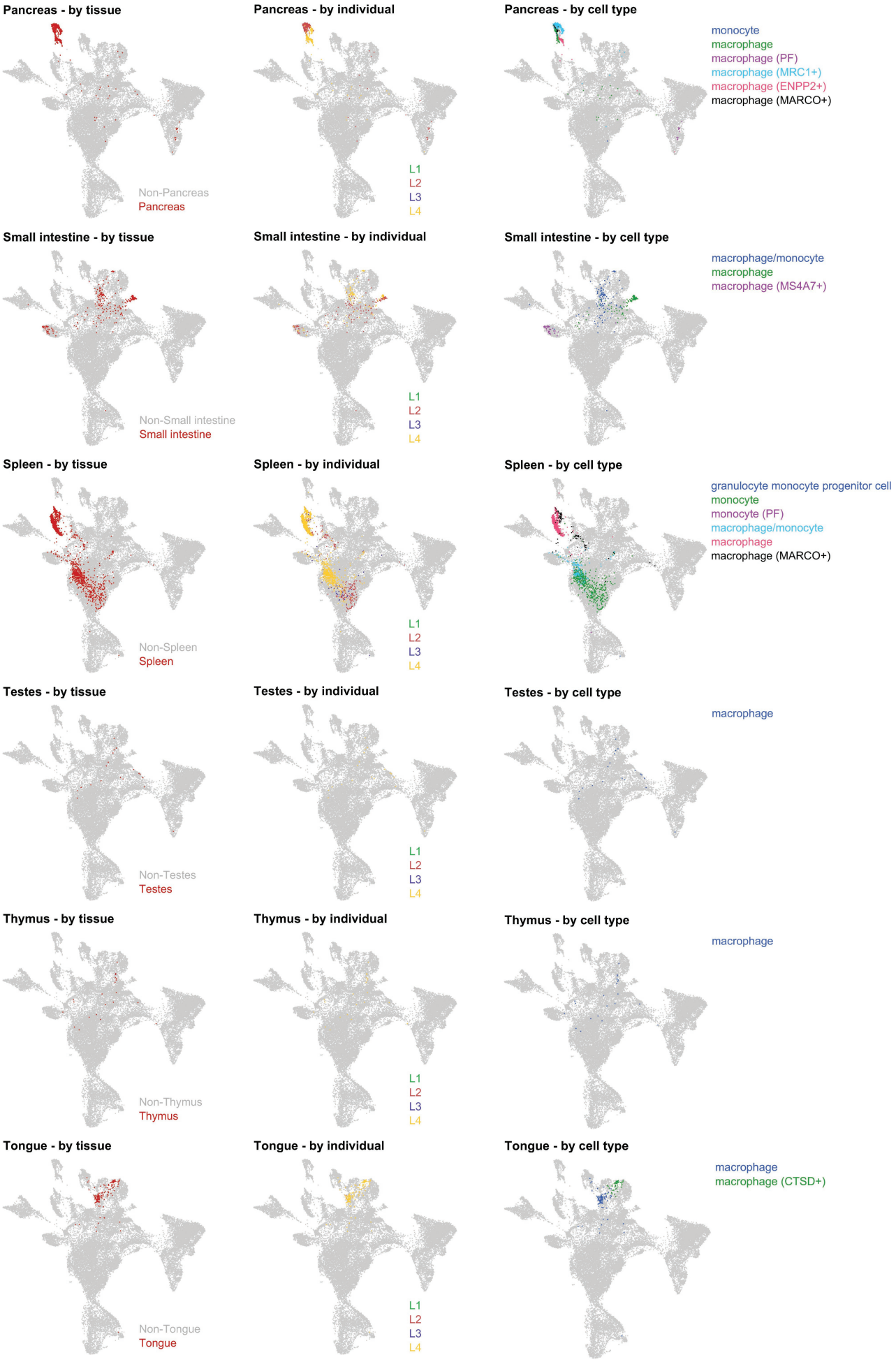
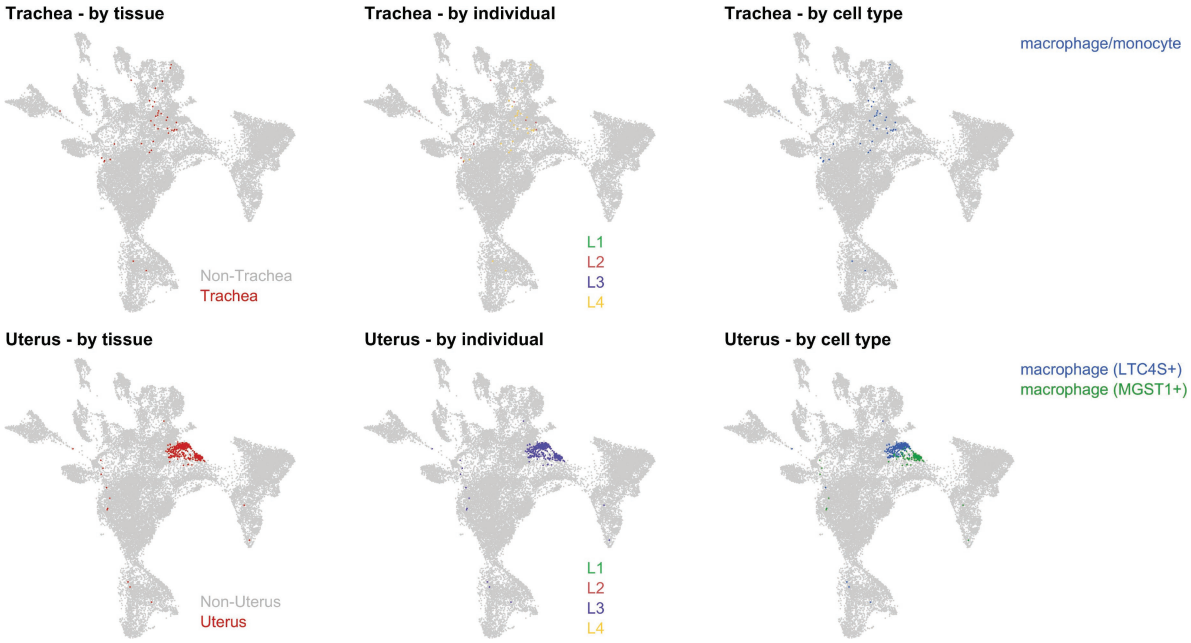


Fig. S10 - continued



Lung
Uterus
a Epithelial cell types
AT2
Tuft cell
Respiratory basal cell
Club cell
Ciliated cell
Uterine metastasis cell
AT1
b
Non-ciliated cell (FXYD4+ MUC16-)
Non-ciliated cells (FXYD4+ MUC16+)
Ciliated cell
Non-ciliated cells (IGFBP1+)
c OXTR
d OXTR
e MUC16
f MUC16
ln(UP10K+1)
0
2
4

Fig. S12

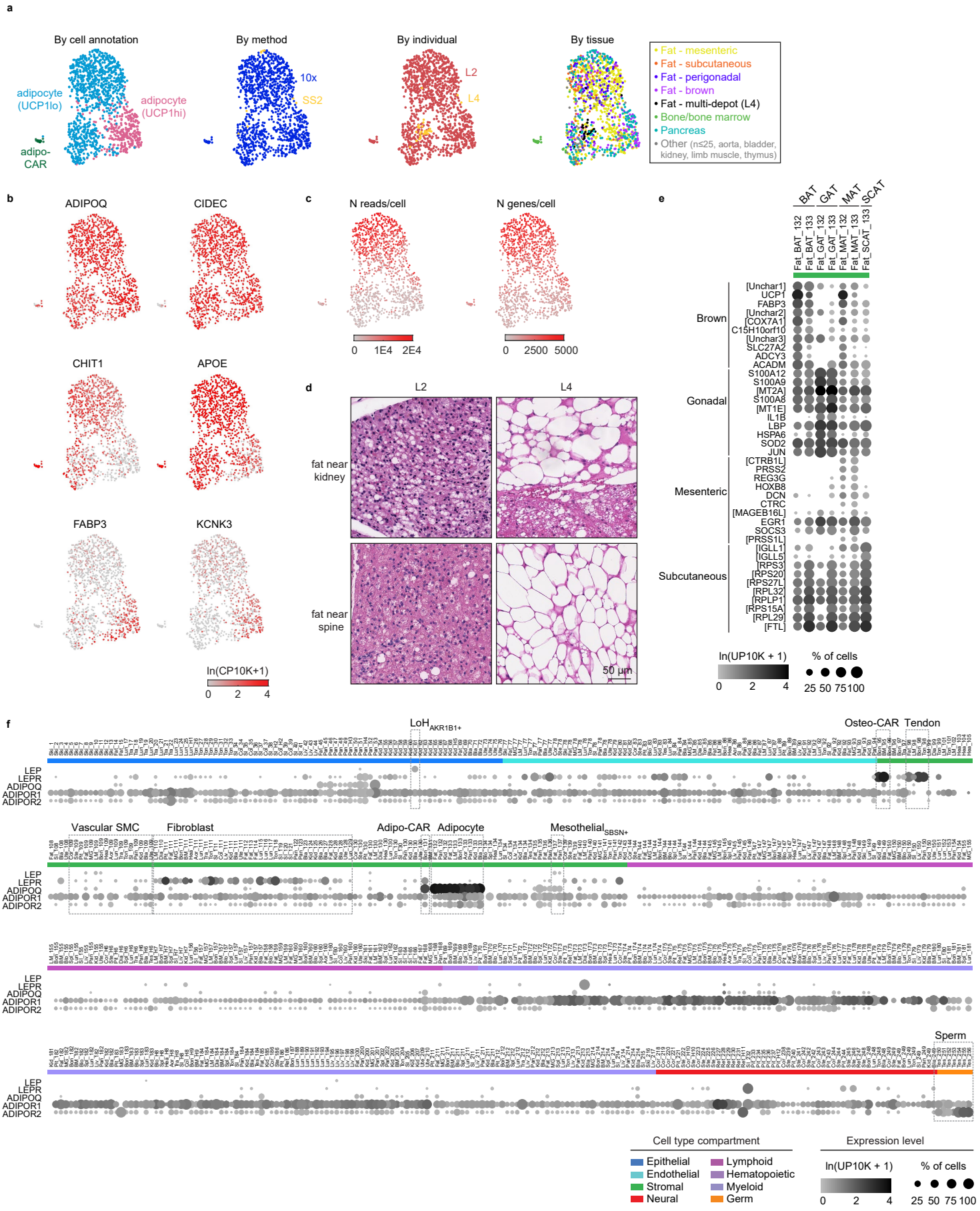
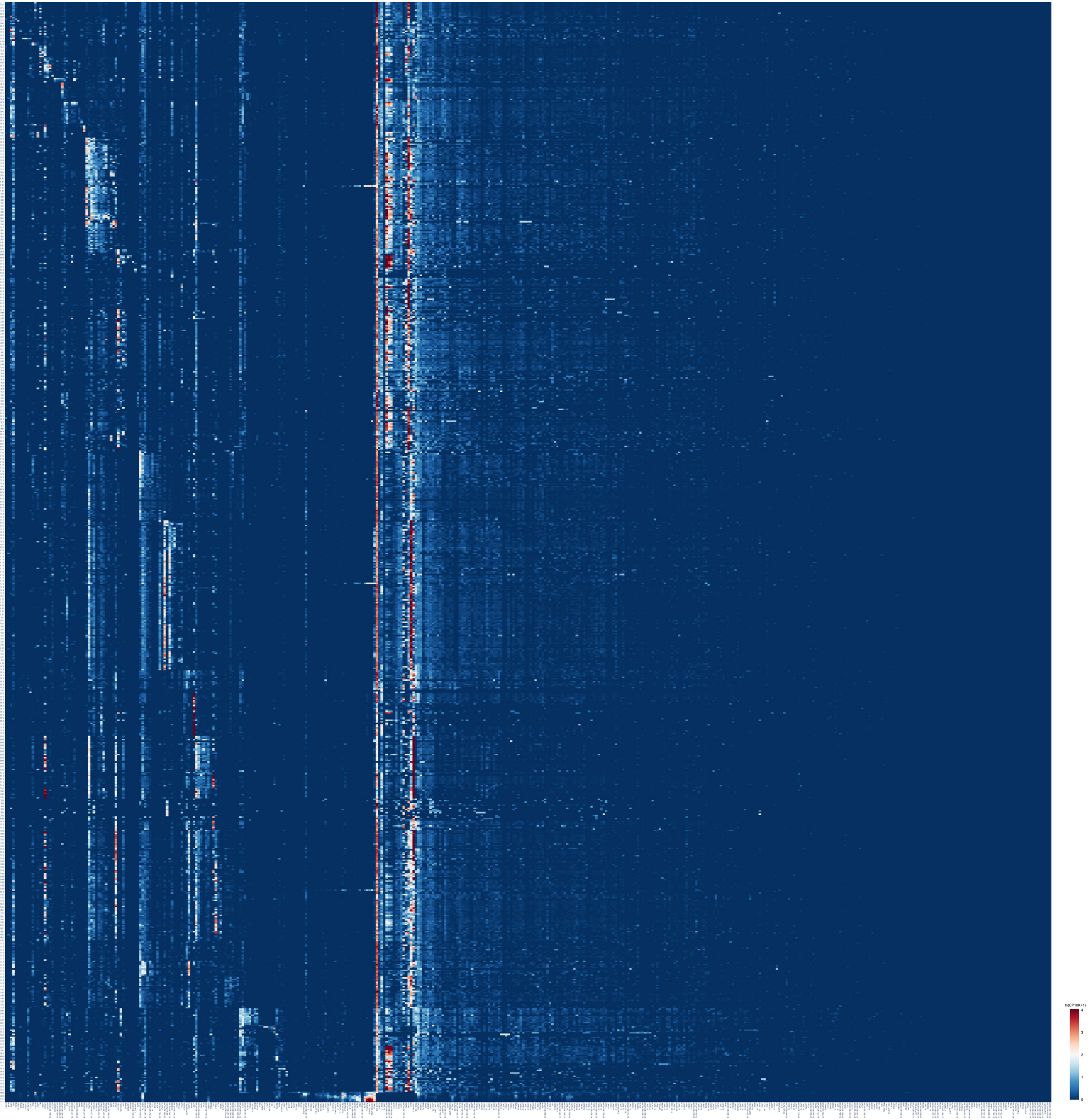

a
By cell annotation
adipocyte (UCP1lo)
adipocyte (UCP1hi)
adipo-CAR
By method
10x
SS2
By individual
L2
L4
By tissue
Fat - mesenteric
Fat - subcutaneous
Fat - perigonadal
Fat - brown
Fat - multi-depot (L4)
Bone/bone marrow
Pancreas
Other (n≤25, aorta, bladder, kidney, limb muscle, thymus)
b
ADIPOQ
CIDEC
CHIT1
APOE
FABP3
KCNK3
ln(CP10K+1)
0 2 4
c
N reads/cell
0 1E4 2E4
N genes/cell
0 2500 5000
d
L2
L4
fat near kidney
fat near spine
50 µm
e
BAT
GAT
MAT
SCAT
Fat_BAT_132
Fat_BAT_133
Fat_GAT_132
Fat_GAT_133
Fat_MAT_132
Fat_MAT_133
Fat_SCAT_133
Brown
[Unchar1]
UCP1
FABP3
[Unchar2]
[COX7A1]
C15H10orf10
[Unchar3]
SLC27A2
ADCY3
ACADM
Gonadal
S100A12
S100A9
[MT2A]
S100A8
[MT1E]
IL1B
LBP
HSPA6
SOD2
JUN
Mesenteric
[CTRB1L]
PRSS2
REG3G
HOXB8
DCN
CTRC
[MAGEB16L]
EGR1
SOCS3
[PRSS1L]
Subcutaneous
[IGLL1]
[IGLL5]
[RPS3]
[RPS20]
[RPS27L]
[RPL32]
[RPLP1]
[RPS15A]
[RPL29]
[FTL]
ln(UP10K + 1)
0 2 4
% of cells
25 50 75 100
f
LoH AKR1B1+
Osteo-CAR
Tendon
Vascular SMC
Fibroblast
Adipo-CAR
Adipocyte
Mesothelial SBSN+
Sperm
LEP
LEPR
ADIPOQ
ADIPOR1
ADIPOR2
Cell type compartment
Epithelial
Endothelial
Stromal
Neural
Lymphoid
Hematopoietic
Myeloid
Germ
Expression level
ln(UP10K + 1)
0 2 4
% of cells
25 50 75 100

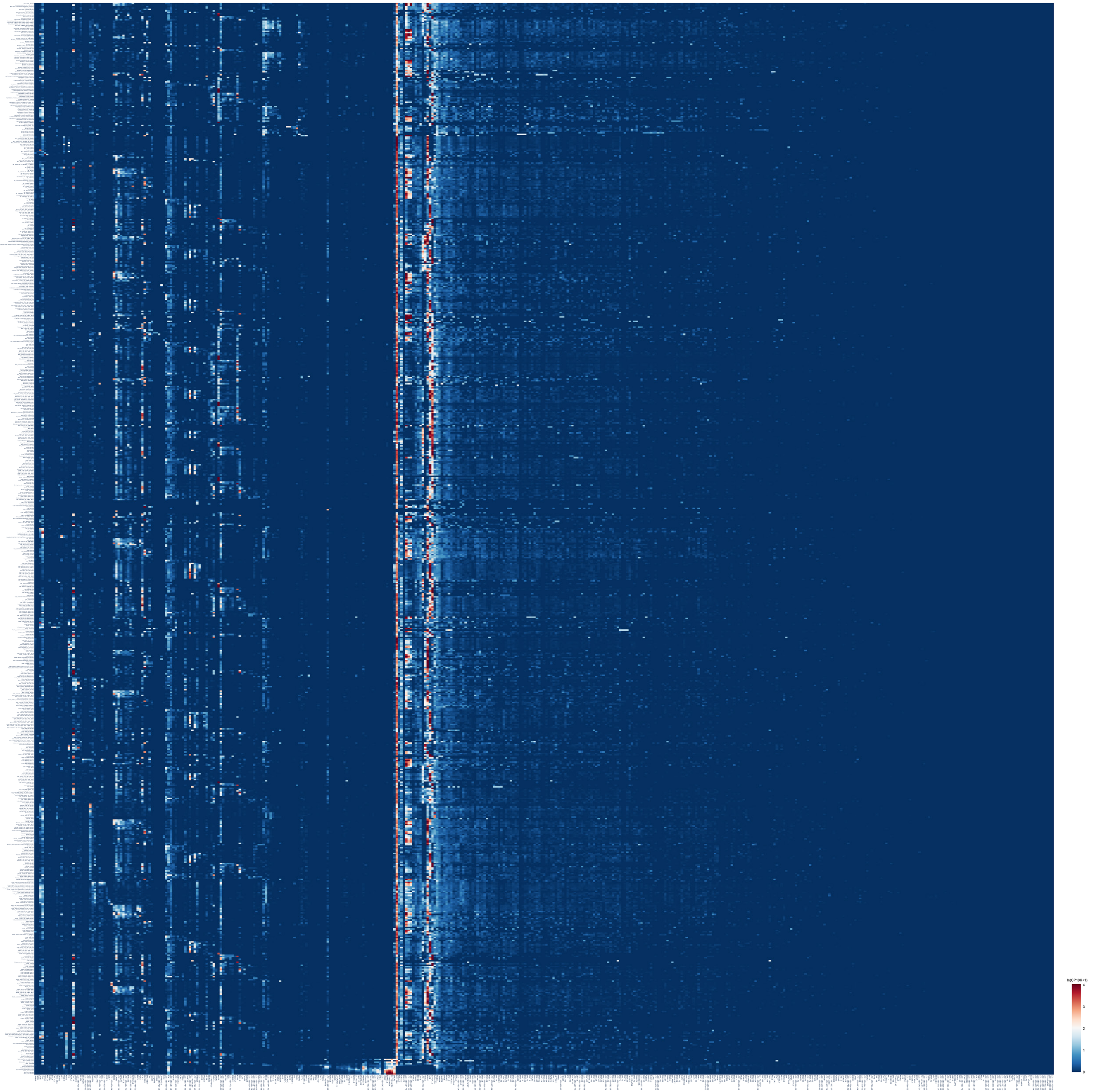
Fig. S13

a. Heatmap of human & lemur-specific genes - ordered by compartment, cell type, tissue



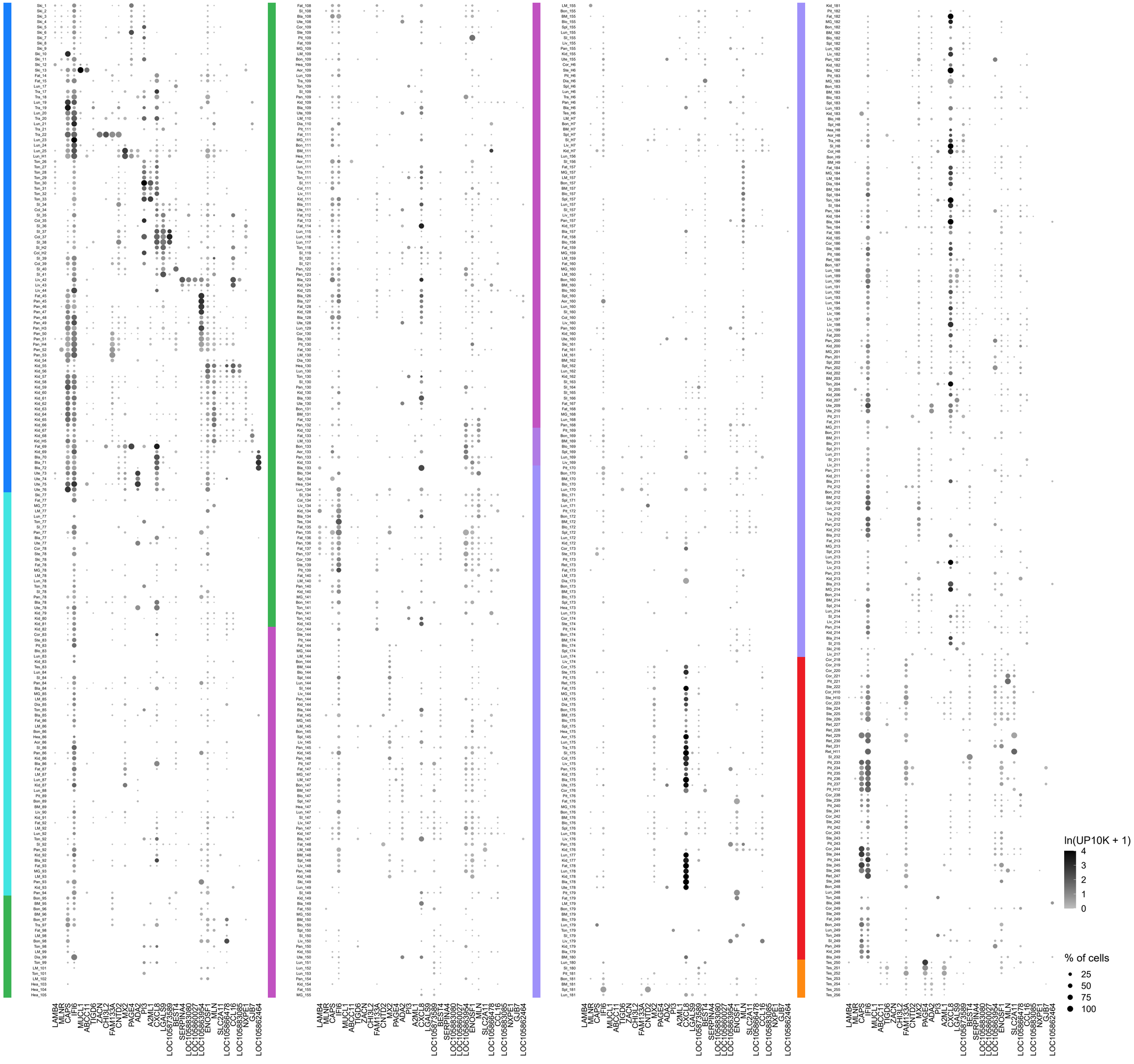
b. Heatmap of human & lemur-specific genes - ordered by tissue, compartment, cell type

Fig. S13 - continued



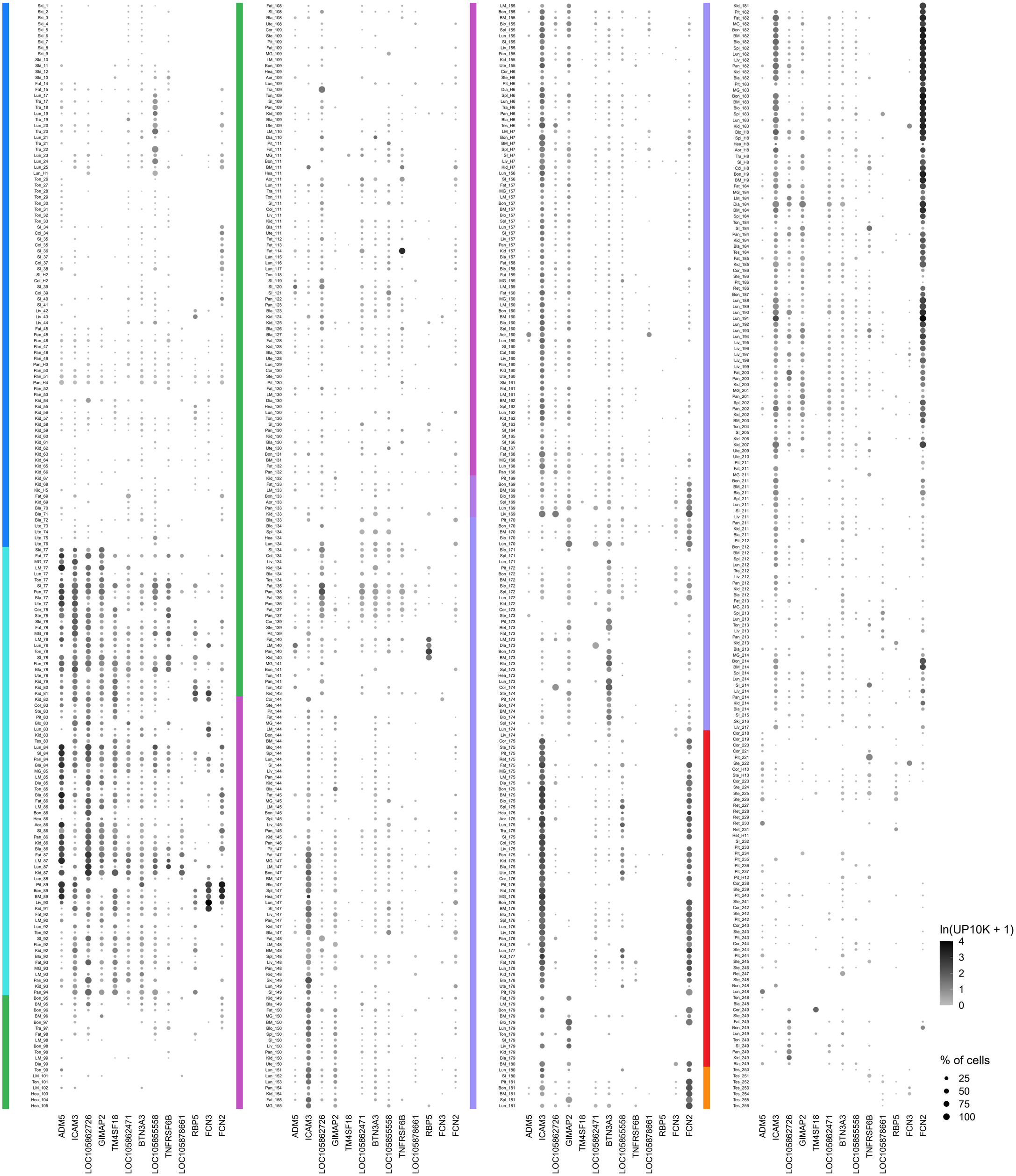
c. Dotplot of human & lemur-specific genes enriched in the epithelial compartment

Fig. S13 - continued



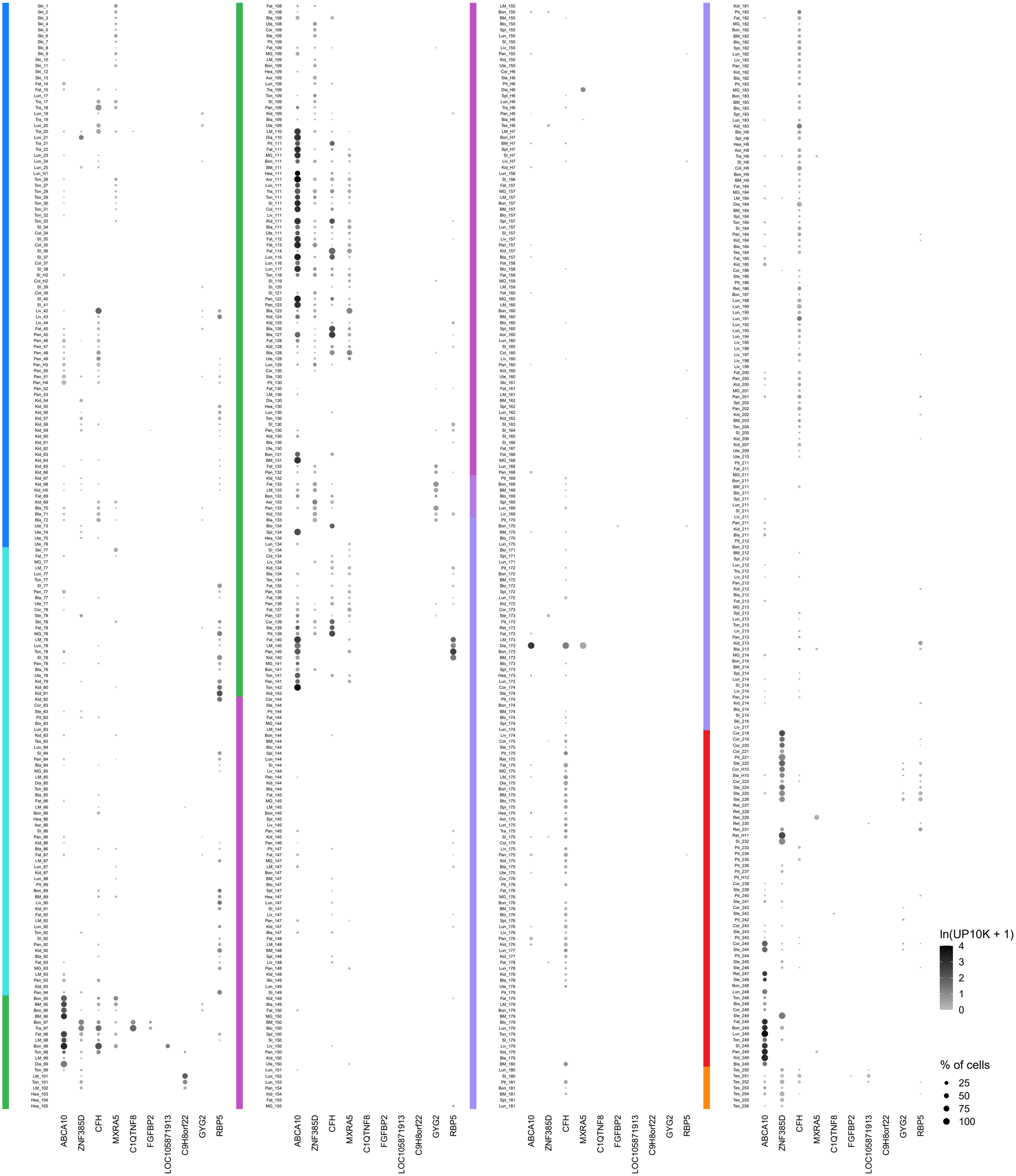
d. Dotplot of human & lemur-specific genes enriched in the endothelial compartment

Fig. S13 - continued



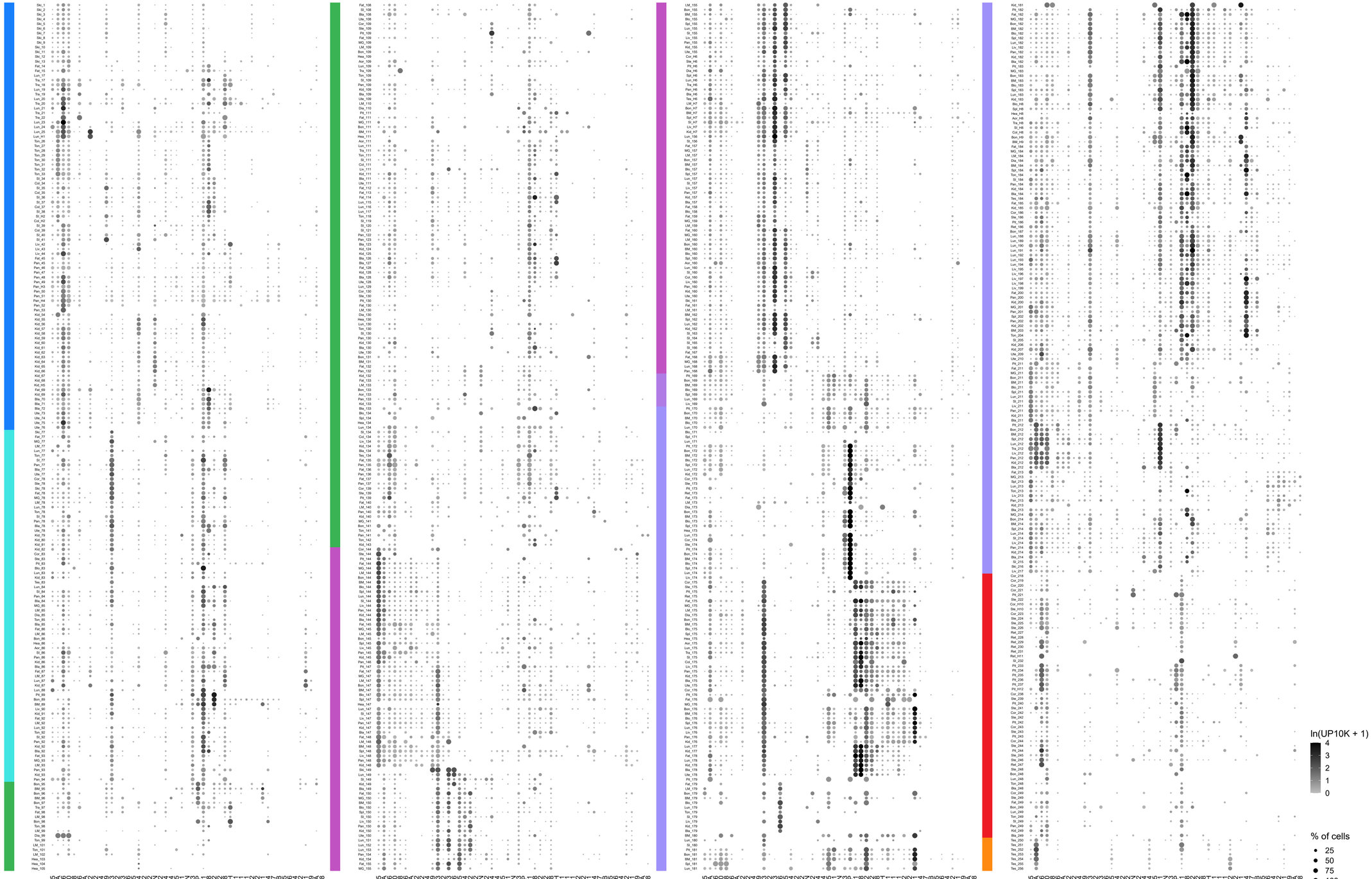
e. Dotplot of human & lemur-specific genes enriched in the stromal compartment

Fig. S13 - continued



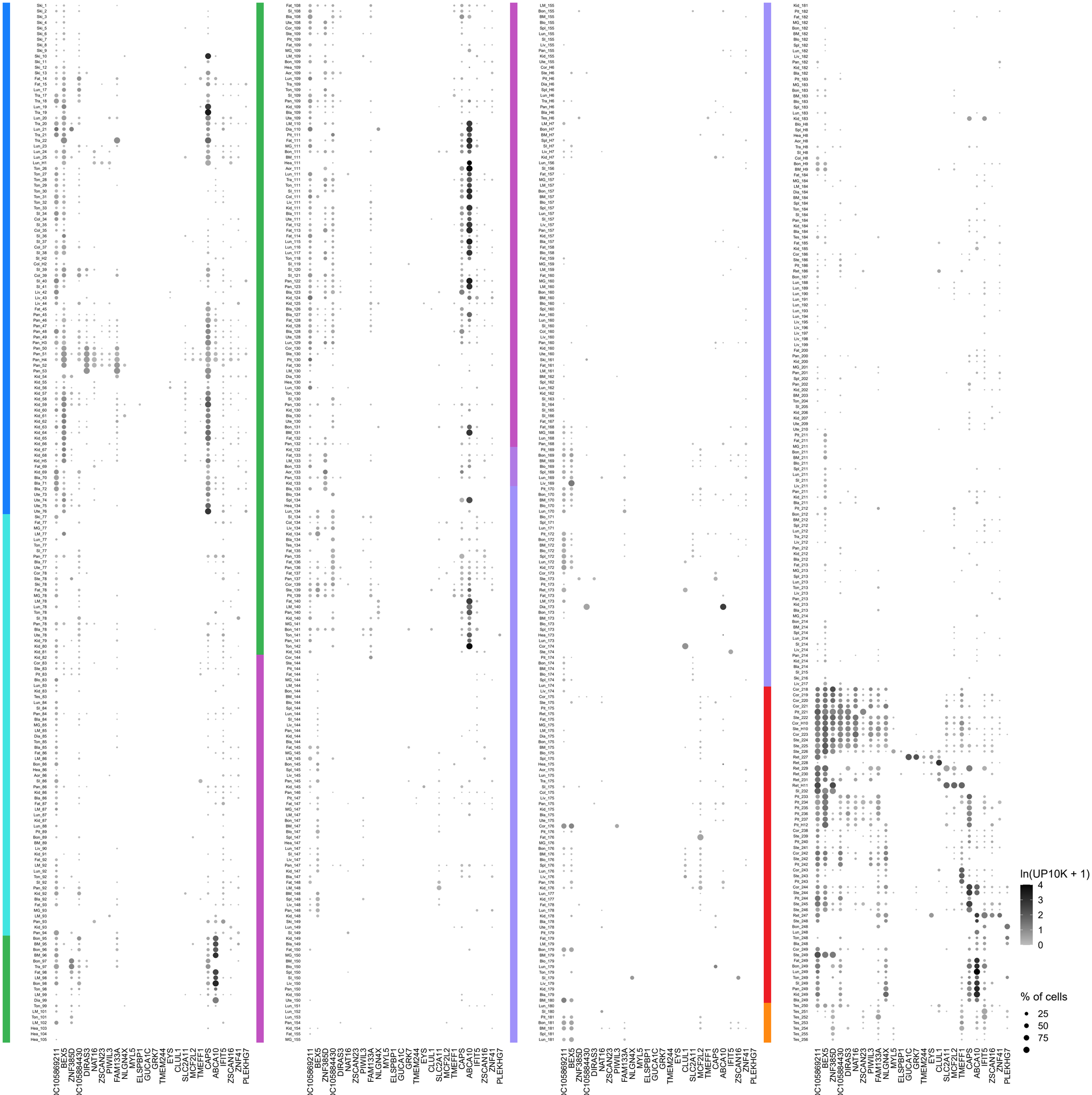
f. Dotplot of human & lemur-specific genes enriched in the immune compartments

Fig. S13 - continued



g. Dotplot of human & lemur-specific genes enriched in the neural compartment

Fig. S13 - continued



h. Dotplot of human & lemur-specific genes enriched and depleted in the germ compartment

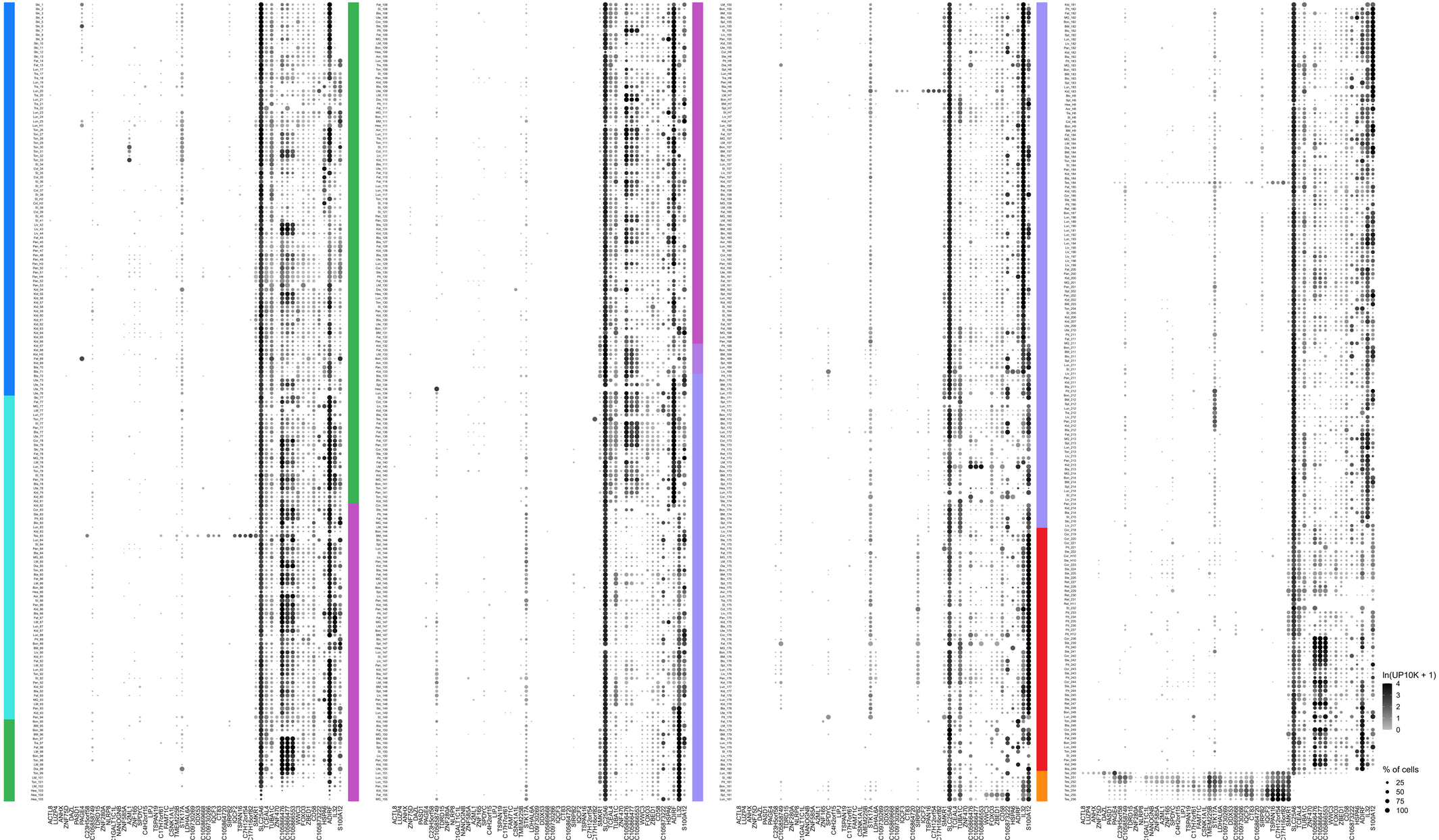



Fig. S13 - continued