

1 **Transcriptomic characterization of 20 organs and tissues from mouse at single cell resolution**
2 **creates a *Tabula Muris***

3
4 The *Tabula Muris* Consortium

5
6 We have created a compendium of single cell transcriptome data from the model
7 organism *Mus musculus* comprising nearly 100,000 cells from 20 organs and tissues.
8 These data reveal many new aspects of cell biology, including gene expression in
9 poorly characterized cell populations and the characterization of new populations of
10 cells in many tissues. Furthermore, they allow for direct and controlled comparison
11 of gene expression in cell types shared between tissues, such as immune cells from
12 distinct anatomical locations. Two distinct technical approaches were used for most
13 organs: one approach, microfluidic droplet-based 3'-end counting^{1,2}, enabled the
14 survey of thousands of cells at relatively low coverage, while the other, FACS-based
15 full length transcript analysis³, enabled characterization of cell types with high
16 sensitivity and coverage. The cumulative resource provides the foundation for a
17 comprehensive resource of transcriptomic cell biology.

18
19 The cell is a fundamental unit of structure and function in biology, and multicellular
20 organisms have evolved a wide variety of different cell types with specialized roles.
21 Although cell types have historically been characterized on the basis of morphology and
22 phenotype, the development of molecular methods has enabled ever more precise
23 definition of their properties, typically by measuring protein or mRNA expression
24 patterns⁴. Technological advances have enabled increasingly greater degrees of
25 multiplexing of these measurements^{5,6}, and it is now possible to use highly parallel
26 sequencing to enumerate nearly every mRNA molecule in a given single cell^{3,7}. This
27 approach has provided many novel insights into cell biology and the composition of
28 organs and tissues from a variety of organisms^{8–17}. However, while these reports provide
29 valuable characterization of individual organs, it is challenging to compare data taken
30 with varying experimental techniques in independent labs from different animals. It
31 therefore remains an open question whether data from individual organs can be
32 synthesized and used as a more general resource for biology.

33
34 Here we report a compendium of cell types from the mouse *Mus musculus*. The
35 compendium is comprised of single cell transcriptome sequence data from 97,029 cells
36 isolated from 20 organs and tissues (Fig. 1). These organs were collected from 3 female
37 and 4 male, C57BL/6 NIA, 3 month old mice (10–15 weeks), whose developmental age is
38 roughly analogous to humans at 20 years of age. We analyzed multiple organs and tissues
39 from the same animal, thereby generating a data set controlled for age, environment and
40 epigenetic effects. This enables the direct comparison of cell type composition between
41 organs as well as comparison of shared cell types across the entire organism. All data,
42 protocols, and analysis scripts from the compendium are shared as a public resource.
43 Gene counts and metadata from all single cells are accessible via Digital Object Identifier
44 (DOI) at 10.6084/m9.figshare.5715040 for the FACS data and
45 10.6084/m9.figshare.5715025 for the microfluidic droplet data. While these data are by
46 no means a complete representation of all mouse organs and cell types, they provide a

47 first draft attempt to create an organism-wide representation of cellular diversity and a
48 comparative framework for future studies using the large variety of murine disease
49 models.

50
51 We developed a specific procedure to collect 20 organs and tissues from the same mouse
52 (see Methods). Briefly, each mouse was anesthetized with 2.5% v/v Avertin, followed by
53 transcardial perfusion with 20 ml PBS. Aorta, bladder, bone marrow, brain (cerebellum,
54 cortex, hippocampus, striatum), colon, diaphragm, fat (brown, gonadal, mesenteric,
55 subcutaneous), heart, kidney, liver, lung, mammary gland, pancreas, skin, skeletal
56 muscle, spleen, thymus, tongue, and trachea were immediately dissected and processed
57 into single cell suspensions (see Extended Data). Organs were dissociated into single
58 cells and either sorted by FACS into 384-well plates (with the exception of liver
59 hepatocytes that were sorted into 96 well plates and heart derived cardiomyocytes that
60 were hand-picked into 96-well plates), and in many cases also loaded into a microfluidic
61 droplet emulsion generating device for cell isolation followed by transcriptome capture
62 (Fig. 1a). Once prepared into sequencing libraries, transcriptomes were sequenced to an
63 average depth of 685,500 reads per cell for the plate data and 7,709 unique molecular
64 identifiers (UMI) per cell for the microfluidic emulsion data. After quality control
65 filtering, 44,879 FACS sorted cells and 55,656 microfluidic droplet processed cells were
66 retained for further analysis. A comparison of the two methods showed differences for
67 each organ in the number of cells analyzed (Fig. 1b,c), reads per cell (Fig. 1d,f) and genes
68 per cell (Fig. 1e,g).

69
70 Using two distinct measurement approaches on the same samples yielded several
71 insights. We compared the methods to understand their relative strengths and
72 weaknesses. Importantly, our results here show that such comparisons can be misleading
73 when performed on a single organ type, as there is substantial variation in performance
74 across organs (Supp. Fig. 1). For example, the lung, trachea and thymus showed nearly
75 twice as many genes detected per cell in the FACS data as compared to the microfluidic
76 data, whereas heart, kidney and marrow show comparable numbers of expressed genes
77 detected by both methods (Supp. Fig. 1-3). Number of genes detected is a fairly crude
78 approach to sensitivity and one must also consider other metrics such as dynamic range
79 for gene expression. The FACS based approach has much higher dynamic range per
80 gene than the microfluidic droplets (Supp. Fig. 4) and enables the analysis of full length
81 transcripts. While the FACS approach generally resulted in higher sensitivity and
82 dynamic range, it also is more laborious and time consuming to perform. The
83 microfluidic droplet approach is faster and enables analysis of larger numbers of cells,
84 albeit at generally lower sensitivity and dynamic range, and with reads only from the 3'
85 end of the transcript. We also expect this study to facilitate the development of
86 computational tools to compare sequencing methods, and our data set will provide a
87 training and validation set for such algorithms. The creation of these tools will be crucial for
88 comparing independent data generated by labs around the world as various tissue atlas
89 studies begin generating results.

90
91 We performed unbiased graph-based clustering of the pooled set of transcriptomes across
92 all organs, and visualized them using tSNE (Fig. 2a). The majority of clusters contain

93 cells from only one organ, but a substantial minority contained cells from many organs.
94 To further dissect these clusters, we separately performed dimensional reduction and
95 clustering on cellular transcriptomes from each individual organ (Fig. 2b). The resulting
96 clusters were manually annotated using *a priori* biological knowledge about cell-type
97 specific gene expression, which confirmed the presence of many specific cell types
98 within organs (Fig. 3). Many of these cell types have not previously been obtained in
99 pure populations and our data provide a wealth of new information on their characteristic
100 gene expression profiles. Initial annotation of each organ and tissue can be found in the
101 extended data, and a detailed discussion each cell types on an organ by organ basis can be
102 found in the supplement. Some unexpected discoveries include: 1) novel immune cell
103 types in the lung, 2) a putative neuroendocrine cell population in the trachea, 3) new
104 differentiated keratinocyte cell types of the tongue, 4) strong sexual dimorphism in
105 hepatocytes with differential gene expression of male pheromones, 5) the suggestions of
106 new roles for genes such as *Neurog3*, *Hhex*, and *Prss53* in the adult pancreas, 6) sexual
107 dimorphism in epithelial and mesenchymal cells of the bladder, and 7) a novel cell
108 population expressing *Chodl* in skeletal muscle.

109
110 In order to better understand the relationships between cell types across organs, we
111 mapped the biologically informed annotations of organ-specific cell types onto the
112 clusters resulting from the unbiased analysis of the complete set of all cell transcriptomes.
113 This analysis revealed that the clusters with cells from many organs generally represent
114 shared cell types common to those organs (Figure 3). For example, B cells from fat, limb
115 muscle, diaphragm, lung, spleen and marrow cluster together, as do T cells from spleen,
116 marrow, lung, limb muscle, fat and thymus. Endothelial cells from fat, heart and lung
117 cluster together, but form a distinct grouping from endothelial cells from mammary,
118 kidney, trachea, limb muscle, aorta, diaphragm and pancreas. There are two clusters
119 containing myeloid cells from limb muscle, brain, diaphragm, aorta, pancreas, kidney,
120 trachea, heart, liver, fat and marrow. Lastly, a cluster containing mesenchymal stem cells
121 from fat, diaphragm, and limb muscle suggests these cells share similar
122 transcriptomes with stromal cells in mammary, trachea, and lung. These findings show
123 that gene-specific manual annotation of cell types is consistent with unbiased clustering
124 based on whole transcriptome profiles, and that cell type identity is strong enough across
125 tissues to enable their unbiased identification at the whole transcriptome level.

126
127 To further investigate a common cell type across organs, we pooled all T-cells and
128 performed an unbiased clustering analysis, revealing 5 clusters (Figure 4) grouped into
129 two distinct sub-groups: clusters 0,4 and clusters 1,2,3. Cluster 0 comprises cells from
130 the Thymus that are undergoing VDJ recombination characterized by the expression of
131 RAG (*Rag1* and *Rag2*) and TdT (*Dnnt*), and includes uncommitted double positive T-
132 cells (*Cd4* and *Cd8a* positive). Cluster 4 contains proliferating T-cells, predominantly
133 from the thymus, and we hypothesize that these are pre-T cells that are expanding after
134 having completed VDJ recombination. In contrast, Clusters 1,2,3 contain mature T-cells.
135 The cells in Cluster 3 are also predominantly from the thymus and show high *Cd5*
136 expression, suggesting that they are undergoing positive selection. Cluster 2 cells are
137 mostly peripheral cells and are most likely activated T-cells expressing the high affinity
138 IL2 receptor (*Il2ra* and *Il2rb*). Interestingly, they also express MHC type II genes (*H2-Aa*

139 and *H2-Ab1*). Finally, Cluster 1 also represents mature T-cells, but primarily from the
140 spleen.

141

142 A key challenge for many single cell studies to date is understanding the potential
143 changes to the transcriptome caused by handling, dissociation and other experimental
144 manipulation. A previous study in limb muscle showed that quiescent satellite cells tend
145 to become activated by dissociation and consequently express immediate early genes
146 among other genes.¹⁸ We found that expression of these dissociation-related markers was
147 also clearly observed in our limb muscle data, as well as in mammary and bladder (Supp.
148 Fig. 6), but that many organs showed little evidence of cellular activation with this panel.
149 Therefore the dissociation-related activation markers found in limb muscle are not
150 universal across all organs. This is not to say that other tissues lack dissociation-related
151 gene expression changes, but that some of the genes involved are specific to a given
152 tissue. It is important to note that the presence of such gene expression changes does not
153 prevent the identification of cell type or the comparison of cell types across organs.

154

155 To understand the extent to which transcription factor (TF) expression determines cell
156 type, we performed a correlation analysis of TF expression¹⁴ across the entire data set for
157 TFs that were significantly enriched ($p < 10^{-5}$, log10 fold change > 0.3) in at least one
158 unique combination of cell type and tissue (e.g., endothelial cells from liver or basal cell
159 of the bladder). (Fig 5A). The off-diagonal elements of the correlation matrix indicate
160 groups of transcription factors with correlated cell type expression. We measured co-
161 expression of these groups of transcription factors and discovered that typically only one
162 of a few cell types used any given combination of two to four TFs. (Supp. Fig. 7) The
163 largest cluster of correlated genes enriched for a broad swath of immune cells across most
164 tissues. Similarly, another cluster of genes enriched for endothelial cells. We also
165 observed groups of TFs specific to several organs; for example within colon, goblet cells
166 were highly enriched for a cluster of TFs that are either known or potential new targets
167 for specification of goblet cells (Fig 5A, Supplementary Figure 7), and we also saw
168 clusters specific to muscle cells, pancreatic cells, neurons, microglia and astrocytes.
169 Another group of genes showed enrichment in several epithelial subtypes of tongue and
170 bladder.

171

172

173 We further analyzed tissue-specific TFs within the endothelial, epithelial and stromal cell
174 types found in several different organs (Fig 5B-D). Within epithelial cells, we found
175 genes grouped by tissue: tongue (*Pax9*), bladder (*Ets1*), skin keratinocyte stem cells
176 (*Lhx2*), and lung (*Aebp1*) (Fig 5B). *Pax9* is enriched in tongue basal cells and is known
177 to be necessary for formation of filiform papillae. *Foxq1* is enriched in bladder basal
178 cells, but its role in bladder basal cell differentiation is heretofore unexplored. Within
179 fibroblast/stromal cells, we found TFs that separated heart from kidney fibroblasts, and
180 specified the stromal cells of lung, trachea, or mammary gland (Fig 5C). Within
181 endothelial cells, we found less strong signatures for tissue specificity; however, liver,
182 brain, kidney, and lung-specific groups of TFs were distinguishable (Fig 5D). These
183 results illustrate how single cell data taken across many organs and tissues can be used in

184 combination to identify which transcriptional regulatory programs are specific to cell
185 types of interest.

186
187 In conclusion, we have created a compendium of single-cell transcriptional
188 measurements across 20 organs and tissues of the mouse. This *Tabula Muris*, or “Mouse
189 Atlas”, has many uses, including the discovery of new putative cell types, the discovery
190 of novel gene expression in known cell types, and the ability to compare cell types across
191 organs. It will also serve as a reference of healthy young adult organs which can be used
192 as a baseline for current and future mouse models of disease. While it is not an
193 exhaustive characterization of all organs of the mouse, it does provide a rich data set of
194 the most highly studied tissues in biology. The *Tabula Muris* provides a framework and
195 description of many of the most populous and important cell populations within the
196 mouse, and represents a foundation for future studies across a multitude of diverse
197 physiological disciplines.

198

199

200 **Acknowledgements**

201

202 This work was funded by the Chan Zuckerberg Biohub, the National Institute on Aging
203 (DP1AG053015, T.W.-C), the Veterans Administration (T.W.-C.), and the Stanford
204 Diabetes Research Center (NIH DK116074, S.K.K.). We thank Sony Biotechnology for
205 making an SH800S instrument available for this project. Some cell sorting/flow
206 cytometry analysis for this project was done on a Sony SH800S instrument in the
207 Stanford Shared FACS Facility. Some fluorescence activated cell sorting (FACS) was
208 done with instruments in the VA Flow Cytometry Core, which is supported by the US
209 Department of Veterans Affairs (VA), Palo Alto Veterans Institute for Research
210 (PAVIR), and the National Institutes of Health (NIH).

211

212 **The *Tabula Muris* Consortium:**

213

214 **Overall Coordination:** Nicholas Schaum^b, Jim Karkanias^r, Norma F Neff^r, Andrew P.
215 May^r, Stephen R. Quake^{o,r,*}, Tony Wyss-Coray^{a,c,d,*}, and Spyros Darmanis^{r,*}

216

217 * to whom correspondence should be addressed: quake@stanford.edu, twc@stanford.edu,
218 spyros.darmanis@czbiohub.org

219

220 **Tissue collection and processing:** Ankit S. Baghel^b, Isaac Bakerman^{b,h,i}, Ishita Bansai^r,
221 Daniela Berdnik^a, Biter Bilen^a, Douglas Brownfield^l, Corey Cain^s, Michelle B. Chen^o,
222 Steven Chen^r, Min Cho^r, Giana Cirolia^r, Stephanie G. Conley^b, Spyros Darmanis^r, Aaron
223 Demers^r, Kubilay Demir^{b,n}, Tessa Divita^r, Haley du Bois^a, Laughing Bear Torrez
224 Dulgeroff^b, Hamid Ebadi^r, Hernan Espinoza^l, Matt Fish^{b,m,n}, Benson M. George^b, Astrid
225 Gillich^l, Foad Green^r, Geraldine Genetiano^r, Xueying Gu^m, Gunsagar S. Gulati^b, Yan
226 Hang^m, Shayan Hosseinzadeh^r, Tal Iram^a, Taichi Isobe^b, Feather Ives^r, Robert Jones^o,
227 Kevin S. Kao^b, Guruswamy Karnam^q, Aaron M. Kershner^b, Bernhard Kiss^{b,x}, William
228 Kong^b, Jonathan Lam^m, Davis P. Lee^d, Song E. Lee^a, Benoit Lehallier^a, Guang Li^j,
229 Qingyun Li^w, Ling Liu^{a,c}, Annie Lo^r, Wan-Jin Lu^{b,l}, Anoop Manjunath^b, Andrew P. May^r,

230 Kaia L. May^r, Oliver L. May^r, Ashley Maynard^r, Marina McKay^r, Marco Mignardi^o,
231 Dullei Min^k, Antoine de Morree^a, Ahmad N. Nabhan^l, Norma F Neff^r, Katharine M. Ng^o,
232 Joseph Noh^b, Rasika Patkar^q, Weng Chuan Peng^m, Lolita Penland^r, Robert Puccinelli^r,
233 Eric J. Rulifson^m, Nicholas Schaum^b, Shaheen S. Sikandar^b, Rahul Sinha^{b,e,f,g}, Rene V
234 Sit^r, Krzysztof Szade^{b,y}, Weilun Tan^r, Cristina Tato^r, Krissie Tellez^m, Kyle J. Travaglini^l,
235 Carolina Tropini^p, Lucas Waldburger^r, Linda J. van Weele^b, Michael N. Wosczyna^a, Jinyi
236 Xiang^b, Soso Xue^o, Justin Youngyunpipatkul^r, Hanadie Yousef^a, Fabio Zanini^o, Macy E.
237 Zardeneta^d, Lu Zhou^w
238

239 **Library preparation and sequencing:** Ishita Bansai^r, Steven Chen^r, Min Cho^r, Giana
240 Cirolia^r, Spyros Darmanis^r, Aaron Demers^r, Tessa Divita^r, Hamid Ebadi^r, Geraldine
241 Genetiano^r, Foad Green^r, Shayan Hosseinzadeh^r, Feather Ives^r, Annie Lo^r, Andrew P.
242 May^r, Ashley Maynard^r, Marina McKay^r, Norma F Neff^r, Lolita Penland^r, Rene V Sit^r,
243 Weilun Tan^r, Lucas Waldburger^r, Justin Youngyunpipatkul^r
244

245 **Computational Data Analysis:** Joshua Batson^r, Olga Botvinnik^r, Paola Castro^r, Spyros
246 Darmanis^r, Joseph L. DeRisi^{r,t}, Jim Karkanias^r, Geoffrey M. Stanley^o, James T. Webber^r
247

248 **Cell Type Annotation:** Ankit S. Baghel^b, Isaac Bakerman^{b,h,i}, Ben A. Barres^w, Joshua
249 Batson^r, Phil Beachy^{b,l,m,n}, Biter Bilen^a, Olga Botvinnik^r, Douglas Brownfield^l, Charles
250 K.F. Chan^v, Michelle B. Chen^o, Michael F. Clarke^b, Spyros Darmanis^r, Kubilay Demir^{b,n},
251 Hamid Ebadi^r, Hernan Espinoza^l, Matt Fish^{b,m,n}, Benson M. George^b, Astrid Gillich^l,
252 Xueying Gu^m, Gunsagar S. Gulati^b, Yan Hang^m, Kerwyn Casey Huang^{o,p,r}, Tal Iram^a,
253 Taichi Isobe^b, Guruswamy Karnam^q, Aaron M. Kershner^b, Seung K. Kim^{m,u}, Bernhard M.
254 Kiss^{b,x}, William Kong^b, Mark A. Krasnow^{l,n}, Christin S. Kuo^{k,l,n}, Jonathan Lam^m, Benoit
255 Lehallier^a, Guang Li^l, Qingyun Li^w, Ling Liu^{a,c}, Wan-Jin Lu^{b,l}, Dullei Min^k, Antoine de
256 Morree^a, Ahmad N. Nabhan^l, Katharine M. Ng^o, Patricia K. Nguyen^{b,h,i,j}, Roel Nusse^{b,m,n},
257 Rasika Patkar^q, Weng Chuan Peng^m, Lolita Penland^r, Thomas A. Rando^{a,c,d}, Eric J.
258 Rulifson^m, Nicholas Schaum^b, Shaheen S. Sikandar^b, Rahul Sinha^{b,e,f,g}, Krzysztof
259 Szade^{b,z}, Krissie Tellez^m, Kyle J. Travaglini^l, Carolina Tropini^p, Linda J. van Weele^b,
260 Bruce M. Wang^q, Kenneth Weinberg^k, Irving L. Weissman^{b,e,f,g}, Michael N. Wosczyna^a,
261 Sean Wu^{b,h,j}, Tony Wyss-Coray^{a,c,d}, Jinyi Xiang^b, Hanadie Yousef^a, Macy E. Zardeneta^d,
262 Lu Zhou^w
263

264 **Writing Group:** Joshua Batson^r, Olga Botvinnik^r, Steven Chen^r, Spyros Darmanis^r,
265 Foad Green^r, Andrew P. May^r, Ashley Maynard^r, Stephen R. Quake^{o,r}, Nicholas Schaum^r,
266 James T. Webber^r, Tony Wyss-Coray^{a,c,d}
267

268 **Principal Investigators:** Ben A. Barres^w, Phil Beachy^{b,l,m,n}, Charles Chan^b, Michael F.
269 Clarke^b, Spyros Darmanis^r, Kerwyn Casey Huang^{o,p,r}, Jim Karkanias^r, Seung K. Kim^{m,u},
270 Mark A. Krasnow^{l,n}, Christin S. Kuo^{k,l,n}, Andrew P. May^r, Norma Neff^r, Roel Nusse^{b,m,n},
271 Patricia K. Nguyen^{b,h,i,j}, Thomas A. Rando^{a,c,d}, Bruce M. Wang^q, Kenneth Weinberg^k,
272 Irving L. Weissman^{b,e,f,g}, Sean Wu^{b,h,j}, Stephen R. Quake^{o,r}, Tony Wyss-Coray^{a,c,d}
273

274 ^aDepartment of Neurology and Neurological Sciences, Stanford University School of
275 Medicine, Stanford, California, USA

- 276 ^bInstitute for Stem Cell Biology and Regenerative Medicine, Stanford University School
277 of Medicine, Stanford, California, USA
278 ^cPaul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine,
279 Stanford, California, USA
280 ^dCenter for Tissue Regeneration, Repair, and Restoration, V.A. Palo Alto Healthcare
281 System, Palo Alto, California, USA
282 ^eDepartment of Pathology, Stanford University School of Medicine, Stanford, California,
283 USA
284 ^fLudwig Center for Cancer Stem Cell Research and Medicine, Stanford University
285 School of Medicine, Stanford, California, USA
286 ^gStanford Cancer Institute, Stanford University School of Medicine, Stanford, California,
287 USA.
288 ^hStanford Cardiovascular Institute, Stanford University School of Medicine, Stanford,
289 California, USA
290 ⁱDepartment of Medicine, Division of Cardiology, Stanford University School of
291 Medicine, Stanford, California, USA
292 ^jDepartment of Medicine, Division of Cardiovascular Medicine, Stanford University,
293 Stanford, California, USA
294 ^kDepartment of Pediatrics, Stanford University school of Medicine, Stanford, California,
295 USA
296 ^lDepartment of Biochemistry, Stanford University School of Medicine, Stanford,
297 California, USA
298 ^mDepartment of Developmental Biology, Stanford University School of Medicine,
299 Stanford, California, USA
300 ⁿHoward Hughes Medical Institute, USA
301 ^oDepartment of Bioengineering, Stanford University, Stanford, California, USA
302 ^pDepartment of Microbiology & Immunology, Stanford University School of Medicine,
303 Stanford, California, USA
304 ^q Department of Medicine and Liver Center, University of California San Francisco, San
305 Francisco, California, USA
306 ^rChan Zuckerberg Biohub, San Francisco, California, USA
307 ^sFlow Cytometry Core, V.A. Palo Alto Healthcare System, Palo Alto, California, USA
308 ^t Department of Biochemistry and Biophysics, University of California San Francisco,
309 San Francisco, California USA
310 ^u Department of Medicine and Stanford Diabetes Research Center, Stanford University,
311 Stanford, California USA
312 ^v Department of Surgery, Division of Plastic and Reconstructive Surgery, Stanford
313 University, Stanford, California USA
314 ^w Department of Neurobiology, Stanford University School of Medicine, Stanford, CA
315 USA
316 ^x Department of Urology, Stanford University School of Medicine, Stanford, California,
317 USA
318 ^y Department of Medical Biotechnology, Faculty of Biophysics, Biochemistry and
319 Biotechnology , Jagiellonian University, Poland
320

321 **Methods**

322

323 **Mice and Tissue Collection**

324 Four 10-15 week old male and four virgin female C57BL/6 mice were shipped from the
325 National Institute on Aging colony at Charles River to the Veterinary Medical Unit
326 (VMU) at the VA Palo Alto (VA). At both locations, mice were housed on a 12-h
327 light/dark cycle, and provided food and water *ad libitum*. The diet at Charles River was
328 NIH-31, and Teklad 2918 at the VA VMU. Littermates were not recorded or tracked, and
329 mice were housed at the VA VMU for no longer than 2 weeks before euthanasia. Prior to
330 tissue collection, mice were placed in sterile collection chambers for 15 minutes to collect
331 fresh fecal pellets. Following anesthetization with 2.5% v/v Avertin, mice were weighed,
332 shaved, and blood drawn via cardiac puncture before transcardial perfusion with 20 ml
333 PBS. Mesenteric adipose tissue (MAT) was then immediately collected to avoid exposure
334 to the liver and pancreas perfusate, which negatively impacts cell sorting. Isolating viable
335 single cells from both pancreas and liver of the same mouse was not possible, therefore, 2
336 males and 2 females were used for each. Whole tissues were then dissected in the
337 following order: colon, spleen, thymus, trachea, tongue, brain, heart, lung, kidney,
338 gonadal adipose tissue (GAT), bladder, diaphragm, skeletal muscle (*tibialis anterior*),
339 skin (dorsal), subcutaneous adipose tissue (SCAT, inguinal pad), mammary glands (fat
340 pads 2, 3, and 4), brown adipose tissue (BAT, interscapular pad), aorta, and bone marrow
341 (spine and limb bones). Following single cell dissociation as described below, cell
342 suspensions were either used for FACS sorting of individual cells into 384-well plates, or
343 for microfluidic droplet library preparation. All animal care and procedures were carried
344 out in accordance with institutional guidelines approved by the VA Palo Alto Committee
345 on Animal Research.

346

347 **Tissue dissociation and sample preparation:** Specific protocols for each tissue are
348 described in the supplement.

349

350

351 **Single Cell Methods**

352

353 **Lysis plate preparation**

354 Lysis plates were created by dispensing 0.4 μ l lysis buffer (0.5U Recombinant RNase
355 Inhibitor (Takara Bio, 2313B), 0.0625% TritonTM X-100 (Sigma, 93443-100ML), 3.125
356 mM dNTP mix (Thermo Fisher, R0193), 3.125 μ M Oligo-dT₃₀VN (IDT,
357 5' AAGCAGTGGTATCAACGCAGAGTACT₃₀VN-3') and 1:600,000 ERCC RNA
358 spike-in mix (Thermo Fisher, 4456740)) into 384-well hard-shell PCR plates (Biorad
359 HSP3901) using a Tempest or Mantis liquid handler (Formulatrix). 96 well lysis plates
360 were also prepared with 4 μ l lysis buffer. All plates were sealed with AlumaSeal CS
361 Films (Sigma-Aldrich Z722634) or Microseal 'F' PCR plate seal (Biorad MSF 1001) and
362 then spun down for 1 minute at 3220 xg and snap frozen on dry ice. Plates were stored at
363 -80°C until used for sorting.

364

365 **FACS sorting**

366 After dissociation, single cells from each tissue were isolated in 384 or 96-well plates via
367 Fluorescence Activated Cell Sorting (FACS). Most tissues were sorted into 384-well
368 plates using SH800S (Sony) sorters. Heart and liver were sorted into 96-well plates and
369 cardiomyocytes were hand-picked into 96-well plates. Skeletal muscle and diaphragm
370 were sorted into 384-well plates on an Aria III (Becton Dickinson) sorter. Most tissues
371 used combinations of fluorescent antibodies to enrich for the presence of known rare cell
372 populations (see tissue section below), but some tissues were simply sorted into plates
373 after removal of dead cells, small and large debris. For all samples, an initial gate was set
374 to exclude small debris, platelets and cell aggregates, and most tissues included a forward
375 scattering gate to select for single cells. One color channel was used in most tissues to
376 stain and exclude dead cells and high prevalence cells that would otherwise dominate the
377 cell population. Color compensation was used whenever necessary. Cells were sorted
378 using the highest purity setting on the sorter (“Single cell” in the case of the SH800s) for
379 all but the rarest cell types, for which the “Ultrapure” setting was used. Sorters were
380 calibrated using FACS buffer before collecting cells from any tissue and after every 8
381 sorted plates to ensure accurate dispensing into plate wells. For a typical sort, a tube
382 containing 1-3ml the pre-stained cell suspension was filtered, vortexed gently and loaded
383 onto the FACS machine. A small number of cells were flowed at low pressure to check
384 cell concentration and amount of debris. Then the pressure was adjusted, flow was
385 paused, the first destination plate was unsealed and loaded, and single-cell sorting started.
386 If a cell suspension was too concentrated, it was diluted it using FACS buffer or 1X PBS.
387 For some cell types (e.g. hepatocytes), 96 well plates were used when it was not possible
388 to sort individual cells accurately into 384 well plates. Immediately after sorting, plates
389 were sealed with a pre-labeled aluminium seal, centrifuged and flash frozen on dry ice.
390 On average, each 384-well plate took around 8 minutes to sort.
391

392 **cDNA synthesis and library preparation**

393 cDNA synthesis was performed using the Smart-seq2 protocol [1,2]. Briefly, 384-well
394 plates containing single-cell lysates were thawed on ice followed by first strand synthesis.
395 0.6 µl of reaction mix (16.7 U/µl SMARTScribe Reverse Transcriptase (Takara Bio,
396 639538), 1.67 U/µl Recombinant RNase Inhibitor (Takara Bio, 2313B), 1.67X First-
397 Strand Buffer (Takara Bio, 639538), 1.67 µM TSO (Exiqon, 5'-
398 AAGCAGTGGTATCAACGCAGACTACATrGrG+G-3'), 8.33 mM DTT (Bioworld,
399 40420001-1), 1.67 M Betaine (Sigma, B0300-5VL), and 10 mM MgCl₂ (Sigma, M1028-
400 10X1ML)) was added to each well using a Tempest liquid handler. Reverse transcription
401 was carried out by incubating wells on a ProFlex 2x384 thermal-cycler (Thermo Fisher)
402 at 42°C for 90 min and stopped by heating at 70°C for 5 min.
403

404 Subsequently, 1.5 µl of PCR mix (1.67X KAPA HiFi HotStart ReadyMix (Kapa
405 Biosystems, KK2602), 0.17 µM IS PCR primer (IDT, 5'-
406 AAGCAGTGGTATCAACGCAGAGT-3'), and 0.038 U/µl Lambda Exonuclease (NEB,
407 M0262L)) was added to each well with a Mantis liquid handler (Formulatrix), and second
408 strand synthesis was performed on a ProFlex 2x384 thermal-cycler by using the
409 following program: 1. 37°C for 30 minutes, 2. 95°C for 3 minutes, 3. 23 cycles of 98°C
410 for 20 seconds, 67°C for 15 seconds, and 72°C for 4 minutes, and 4. 72°C for 5 minutes.
411

412 The amplified product was diluted with a ratio of 1 part cDNA to 10 parts 10mM Tris-
413 HCl (Thermo Fisher, 15568025), and concentrations were measured with a dye-
414 fluorescence assay (Quant-iT dsDNA High Sensitivity kit; Thermo Fisher, Q33120) on a
415 SpectraMax i3x microplate reader (Molecular Devices). Sample plates were selected for
416 downstream processing if the mean concentration of blanks - ERCC-containing, non-cell
417 wells - was greater than 0 ng/ μ l, and, after linear regression of the values obtained from
418 the Quant-iT dsDNA standard curve, the R² value was greater than 0.98. Sample wells
419 were then selected if their cDNA concentrations were at least one standard deviation
420 greater than the mean concentration of the blanks. These wells were reformatted to a new
421 384-well plate at a concentration of 0.3 ng/ μ l and final volume of 0.4 μ l using an Echo
422 550 acoustic liquid dispenser (Labcyte).

423
424 Illumina sequencing libraries were prepared as described in Darmanis et al. 2015.¹³
425 Briefly, fragmentation was carried out on double-stranded cDNA using the Nextera XT
426 Library Sample Preparation kit (Illumina, FC-131-1096). Each well was mixed with 0.8
427 μ l Nextera fragmentation DNA buffer (Illumina) and 0.4 μ l Tn5 enzyme (Illumina), then
428 incubated at 55°C for 10 min. The reaction was stopped by adding 0.4 μ l “Neutralize
429 Fragment Buffer” (Illumina) and spinning at room temperature in a centrifuge at 3220xg
430 for 5 min. Indexing PCR reactions were performed by adding 0.4 μ l of 5 μ M i5 indexing
431 primer, 0.4 μ l of 5 μ M i7 indexing primer, and 1.2 μ l of Nextera NPM mix (Illumina).
432 PCR amplification was carried out on a ProFlex 2x384 thermal cycler using the following
433 program: 1. 72°C for 3 minutes, 2. 95°C for 30 seconds, 3. 12 cycles of 95°C for 10
434 seconds, 55°C for 30 seconds, and 72°C for 1 minute, and 4. 72°C for 5 minutes.
435

436 **Library pooling, quality control, and sequencing**

437 Following library preparation, wells of each library plate were pooled using a
438 Mosquito liquid handler (TTP Labtech). Pooling was followed by two purifications using
439 0.7x AMPure beads (Fisher, A63881). Library quality was assessed using capillary
440 electrophoresis on a Fragment Analyzer (AATI), and libraries were quantified by qPCR
441 (Kapa Biosystems, KK4923) on a CFX96 Touch Real-Time PCR Detection System
442 (Biorad). Plate pools were normalized to 2 nM and equal volumes from 10 or 20 plates
443 were mixed together to make the sequencing sample pool. PhiX control library was
444 spiked in at 0.2% before sequencing.
445

446 **Sequencing libraries from 384-well and 96-well plates**

447 Libraries were sequenced on the NovaSeq 6000 Sequencing System (Illumina) using 2 x
448 100bp paired-end reads and 2 x 8bp or 2 x 12bp index reads with either a 200 or 300-
449 cycle kit (Illumina, 20012861 or 20012860).
450
451

452 **Microfluidic droplet single cell analysis**

453 Single cells were captured in droplet emulsions using GemCode Single-Cell Instrument
454 (10x Genomics, Pleasanton, CA, USA) and SC RNA-seq libraries were constructed as
455 per 10X Genomics protocol using GemCode Single-Cell 3' Gel Bead and Library V2
456 Kit. Briefly, single cell suspensions were examined using an inverted microscope and if
457 sample quality was deemed satisfactory, sample was diluted in PBS/2% FBS to achieve a

458 target concentration of 1000 cells/ μ l. If cell suspensions contained cell aggregate or
459 debris, two additional washes in PBS/2% FBS at 300 x g for 5min at 4C were
460 performed. Cell concentration was measured either with a Moxi GO II (Orflo
461 Technologies) or a hemocytometer. Cells were loaded in each channel with a target
462 output of 5,000 cells per sample. All reactions were performed in Biorad C1000 Touch
463 Thermal cycler with 96-Deep Well Reaction Module. 12 cycles were used for cDNA
464 amplification and sample index PCR. Amplified cDNA and final library were evaluated
465 on a Fragment Analyzer using a High Sensitivity NGS Analysis Kit (Advanced
466 Analytical). 10x cDNA libraries were quantitated for average fragment length using a 12
467 or 96 capillary Fragment Analyzer (Advanced Analytical) and by qPCR with Kapa
468 Library Quantification kit for Illumina. Each library was diluted to 2nM and equal
469 volumes of 16 libraries were pooled to fill each NovaSeq sequencing run. Pools were
470 sequenced with 100 cycle run kits with 26 bases Read 1, 8 bases Index 1 and 90 bases
471 Read 2 (Illumina 20012862). PhiX control library was spiked in at 0.2 to 1%. Libraries
472 were sequenced on the NovaSeq 6000 Sequencing System (Illumina)
473

474 **Data Processing**

475 Sequences from the NovaSeq were demultiplexed using bcl2fastq version 2.19.0.316.

476 Reads were aligned using to a copy of the mm10 genome augmented with ERCC

477 sequences, using STAR version 2.5.2b with the following options:

```
478 --outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhang  
479 Min 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilt  
480 erMismatchNoverLmax 0.04 --alignIntronMin 20 --alignIntronMax 100000  
481 0 --alignMatesGapMax 1000000 --outSAMstrandField intronMotif --outSA  
482 Mtype BAM Unsorted --outSAMattributes NH HI NM MD --genomeLoad LoadA  
483 ndKeep --outReadsUnmapped Fastx --readFilesCommand zcat
```

485 Gene counts were produced using HTSEQ version 0.6.1p1 with default parameters
486 except stranded was set to false and mode was set to intersection-nonempty.

488 Sequences from the microfluidic platform were demultiplexed and aligned using
489 CellRanger with default parameters.

492 **Clustering**

493 Standard procedures for filtering, variable gene selection, dimensionality reduction, and
494 clustering were performed using the Seurat package. For each tissue and each sequencing
495 method (FACS and microfluidic), the following steps were performed. (The parameters
496 used were tuned on a per-tissue basis. Values appear in the supplement.)

- 498 1. Genes appearing in fewer than 5 cells were excluded.
- 499 2. Cells with fewer than 500 genes were excluded. Cells with fewer than 50,000
500 reads (FACS) or 1000 UMI (microfluidic) were excluded. (In some organs, cells
501 with more than 2 million reads were also excluded as a conservative measure to
502 avoid doublets.)
- 503 3. Counts were log-normalized ($\log(1 + \text{counts per N})$), then scaled by linear
504 regression against the number of reads (or UMIs), the percent of reads mapping to
505 Rn45s, and the percent of reads to ribosomal genes.

- 506 4. Variable genes were selected using a threshold for dispersion ($\log \text{of}$
507 $\text{variance}/\text{mean}$). The distribution of dispersion in each expression bin was
508 computed, and variable genes were those with a z-score of at least 0.5-1..
509 5. The variable genes were projected onto a low-dimensional subspace using
510 principal component analysis. The number of principal components was selected
511 based on inspection of the plot of variance explained.
512 6. A shared-nearest-neighbors graph was constructed based on the Euclidean
513 distance metric in the low-dimensional subspace. (The SNN is a weighted graph
514 where w_{ij} is the Jaccard similarity of the k-neighborhood of i with the k-
515 neighborhood of j, where $k = 30$.) Cells were clustered using the Louvain method
516 to maximize modularity.
517 7. Cells were visualized using a 2-dimensional t-distributed Stochastic Neighbor
518 Embedding on the same distance metric.
519 8. Cell types were assigned to each cluster using the abundance of known marker
520 genes. (Plots showing the expression of the markers for each tissue appear in the
521 supplement).
522 9. When clusters appeared to be mixtures of cell types, they were refined either by
523 increasing the resolution parameter for Louvain clustering or subsetting the data
524 and rerunning steps 3-7.

525
526 A similar analysis was done globally for all FACS processed cells and for all microfluidic
527 processed cells.

528 To attempt to remove the effect of technical variation from the gene expression matrix, it
529 is common practice to regress out statistics associated with quality, like the number of
530 reads or the percent of ribosomal or mitochondrial RNA. If those statistics differ
531 systematically between cell types or tissues, this can have the unwanted effect of
532 compressing the true variation between those groups. For example, the percent of
533 ribosomal RNA varies from 1% in the Liver and Pancreas to 5% in the Spleen and
534 Tongue. Hence we regress out these factors separately in the analysis of each tissue.
535

536 For the "dynamic range" comparison in Supplementary Figure 4, the gene expression was
537 scaled to log of reads per million for FACS and log of UMI per 10,000 for microfluidic
538 emulsions. Genes with nonzero expression were rank ordered and plotted.

544 **References**

545

- 546 1. Thorsen, T., Roberts, R. W., Arnold, F. H. & Quake, S. R. Dynamic pattern
547 formation in a vesicle-generating microfluidic device. *Phys. Rev. Lett.* **86**, (2001).
- 548 2. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of
549 individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- 550 3. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and
551 individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
- 552 4. Alberts, B. *et al.* *Molecular biology of the cell*. (Garland Science, 2002).
- 553 5. Guo, G. *et al.* Resolution of Cell Fate Decisions Revealed by Single-Cell Gene
554 Expression Analysis from Zygote to Blastocyst. *Dev. Cell* **18**, 675–685 (2010).
- 555 6. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human
556 colon tumors. *Nat. Biotechnol.* **29**, (2011).
- 557 7. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods.
558 (2013). doi:10.1038/nmeth.2694
- 559 8. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium
560 using single-cell RNA-seq. *Nature* **509**, (2014).
- 561 9. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional
562 Signatures of Aging and Somatic Mutation Patterns. *Cell* **171**, (2017).
- 563 10. Halpern, K. B. *et al.* Single-cell spatial reconstruction reveals global division of
564 labour in the mammalian liver. *Nature* **542**, 352–356 (2017).
- 565 11. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature*
566 **551**, 333–339 (2017).
- 567 12. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood
568 dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
- 569 13. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single
570 cell level. *Proc. Natl. Acad. Sci. U. S. A.* **112**, (2015).
- 571 14. Gokce, O. *et al.* Cellular Taxonomy of the Mouse Striatum as Revealed by Single-
572 Cell RNA-Seq. *Cell Rep.* **16**, (2016).
- 573 15. Usoskin, D. *et al.* Unbiased classification of sensory neuron types by large-scale
574 single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2014).
- 575 16. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus
576 revealed by single-cell RNA-seq. *Science* **347**, 1138–42 (2015).
- 577 17. Li, H. *et al.* Classifying Drosophila Olfactory Projection Neuron Subtypes by
578 Single-Cell RNA Sequencing. *Cell* **171**, 1206–1220.e22 (2017).
- 579 18. van den Brink, S. C. *et al.* Single-cell sequencing reveals dissociation-induced
580 gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).

581

582

583

584 Figure captions

585

586 **Figure 1.** a) 19 tissues from 4 male and 3 female mice were analyzed. After tissue
587 dissociation, cells were either sorted by FACS or manipulated in microfluidic emulsions,
588 after which they were lysed and their transcriptomes amplified, followed by sequencing,
589 read mapping, and data analysis. b) Barplot showing number of sequenced cells prepared
590 by FACS sorting from each tissue (n=20). c) Barplot showing number of sequenced cells
591 prepared by microfluidic emulsion from each tissue (n=12). d) Histogram of number of
592 reads per cell for each tissue from FACS sorted cells. e) Histogram of number of genes
593 detected per cell for each tissue from FACS sorted cells. f) Histogram of number of
594 unique molecular identifiers (UMI) sequenced per cell for each tissue from cells prepared
595 by microfluidic emulsion. g) Histogram of number of genes detected per cell for each
596 tissue for cells prepared by microfluidic emulsion.

597

598 **Figure 2.** a) Dimensionally reduced tSNE plot of all cells sorted by FACS color coded
599 by tissue of origin. b) Dimensionally reduced tSNE plots for each tissue of cells sorted by
600 FACS. Color coding indicates distinct clusters. c) Barplots of manually annotated cell
601 types based on differential gene expression across all tissues.

602

603 **Figure 3.** Comparison of cell type determination as done by unbiased whole
604 transcriptome comparison versus manual annotation by organ-specific experts. The x-
605 axis represents clusters from Figure 2a with multiple organs contributing, while the y-
606 axis represents manual expert annotation of cell types in an organ-specific fashion based
607 on the data in Figure 2b and 2c. The unbiased method discovers relationships between
608 similar cell types found in different organs (highlighted regions); in particular it groups T
609 cells from different organs into a single cluster, B cells from different organs into a
610 different single cluster, and endothelial cells from different organs into a single cluster.

611

612 **Figure 4.** Analysis of all T cells sorted by FACS. a) Dimensionally reduced tSNE plot of
613 all T cells colored by cluster membership. Five clusters were identified. b) Dotchart
614 showing level of expression (color scale) and number of expressing cells (point diameter)
615 within each cluster of T cells. c) Barplot showing fraction of T cells originating from Fat,
616 Lung, Marrow, Muscle, Spleen or Thymus for each of the T cell clusters. d) Barplot
617 showing fraction of Cd4⁺/ Cd8⁺/ Cd4⁺Cd8⁺ / Cd4⁻Cd8⁻ T cells for each of the T cell
618 clusters.

619

620 **Figure 5.** Transcription factor expression analysis. A. Gene-gene correlation (correlogram) of top
621 cell type-tissue specific transcription factors (TFs) (selected by minimum differential
622 expression of 0.3 and p-value of 10⁻⁶; immediate early genes excluded). Columns are
623 colored by the cell ontology (left column) and tissue (right column) of the cell
624 type+tissue with the highest expression of that gene. Insets show clusters of TFs enriched
625 in selected cell types and tissues. B. Correlogram of top tissue-specific TFs for epithelial
626 cells (cells ontologies containing “epithelial”, “basal”, “keratinocyte”, or “epidermis”).
627 Row colors correspond to tissue (right) and cell ontology (left) of most-enriched cell
type. tSNE plots are calculated using all variable genes and show tissue origin (right) and

628 gene expression of select TFs (bottom). C. Correlogram and tSNE of tissue-specific TFs
629 within stromal cells (cells ontologies containing “stromal”, “fibroblast”, or
630 “mesenchymal”). D. Correlogram and tSNE of tissue-specific TFs within endothelial
631 cells.

632

633 **Supplementary Figure Captions**

634

635 **Supplementary Figure 1** Histograms of number of genes detected across tissues (Red =
636 FACS sorted cells, Cyan = cells processed in microfluidic emulsion)

637

638 **Supplementary Figure 2** Plots across three cell types which compare number of cells
639 expressing a given gene for measurements made by FACS (x axis) compared to
640 microfluidic emulsion (y axis). Each data point is a gene and the axes represent fraction
641 of cells which express that gene.

642

643 **Supplementary Figure 3** Correlation of number of genes detected in more cells across
644 all common cell types. X-axis represents number of genes present in cell types measured
645 by FACS, whereas Y-axis represents number of genes present in the same cell types
646 measured by microfluidic emulsion. Most but not all of the cell types have more genes
647 measured by FACS.

648

649 **Supplementary Figure 4** Plots showing examples of dynamic range for the two
650 measurements approaches; FACS generally has many more reads per gene.

651

652 **Supplementary Figure 5** a) Dimensionally reduced tSNE plot of all cells manipulated
653 by microfluidic emulsions color coded by tissue of origin. b) Dimensionally reduced
654 tSNE plots for each tissue of cells sorted by microfluidic emulsions. Color coding
655 indicates distinct clusters. c) Barplots of manually annotated cell types based on
656 differential gene expression across all tissues.

657

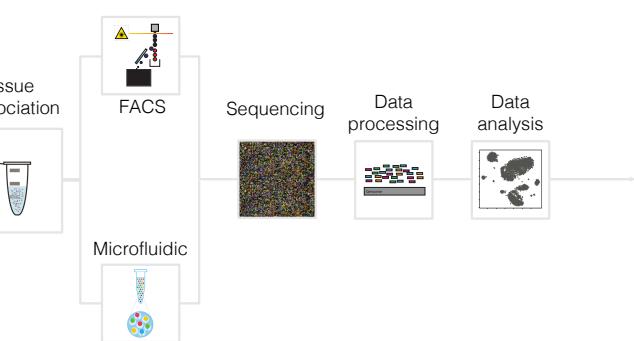
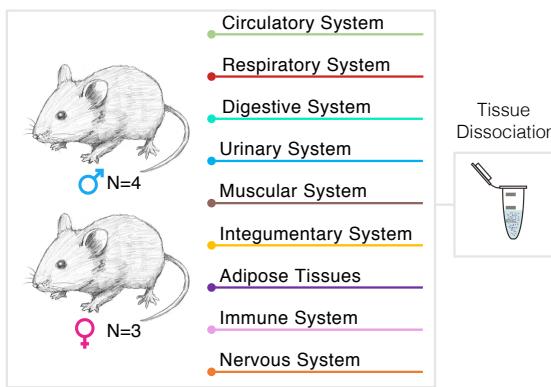
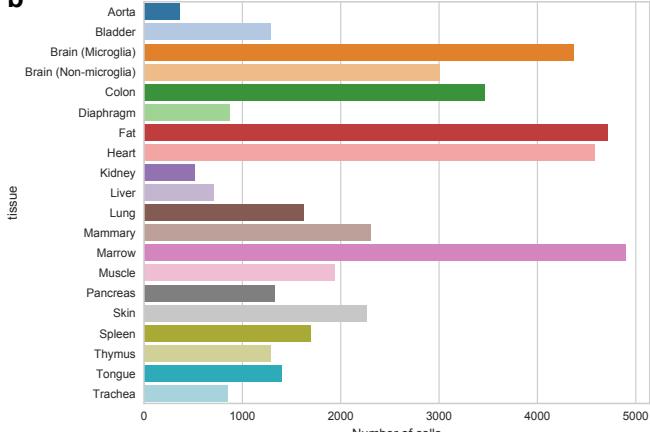
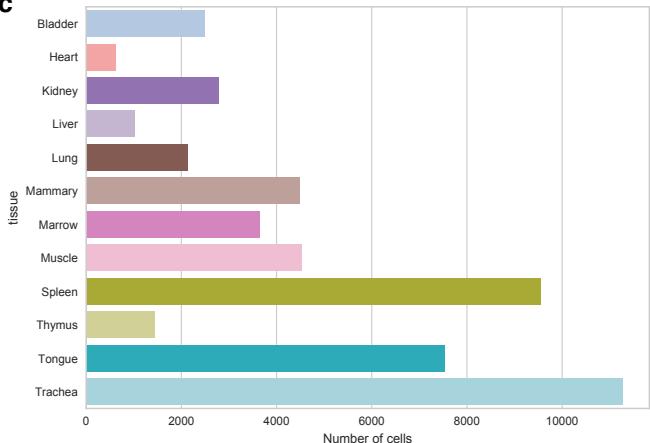
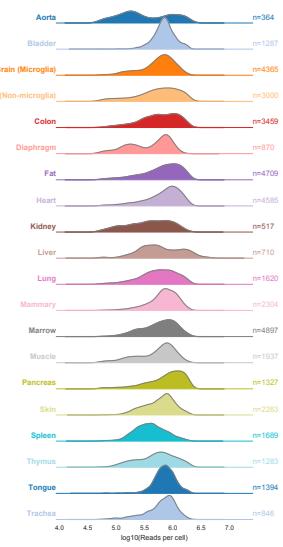
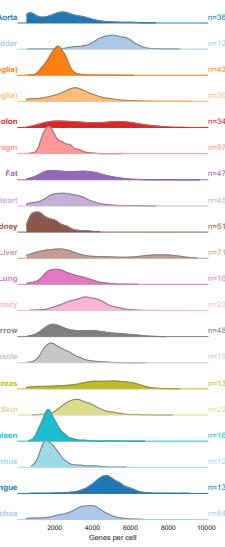
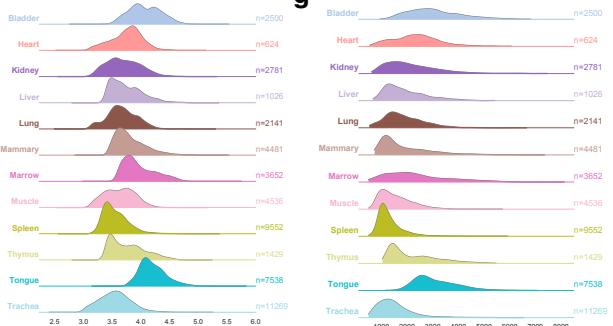
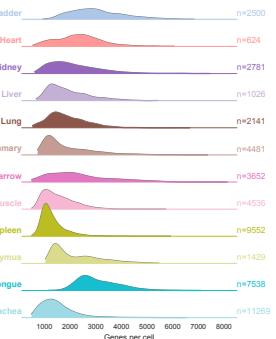
658 **Supplementary Figure 6** Analysis of dissociation induced gene expression scores
659 across tissues.

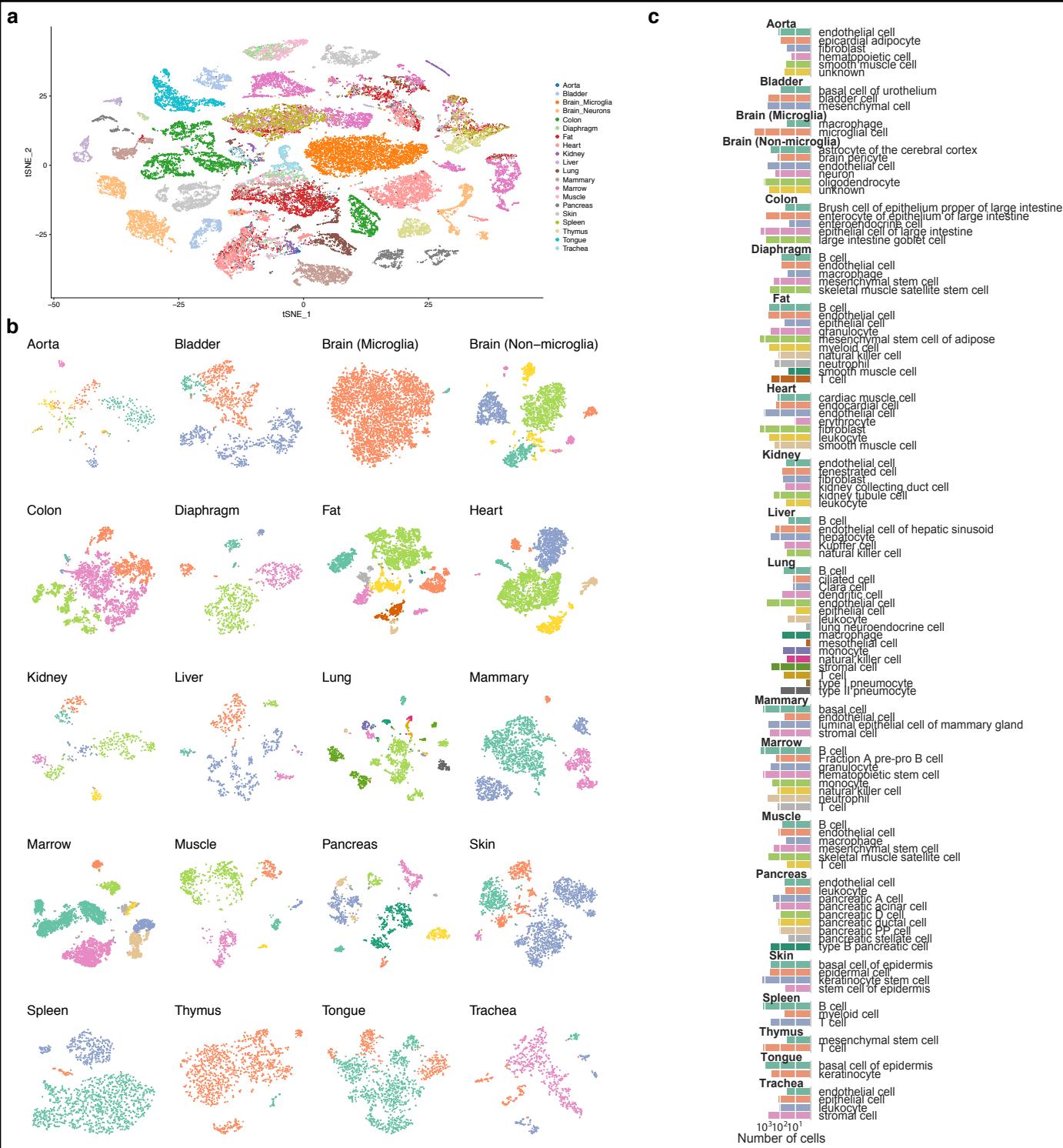
660

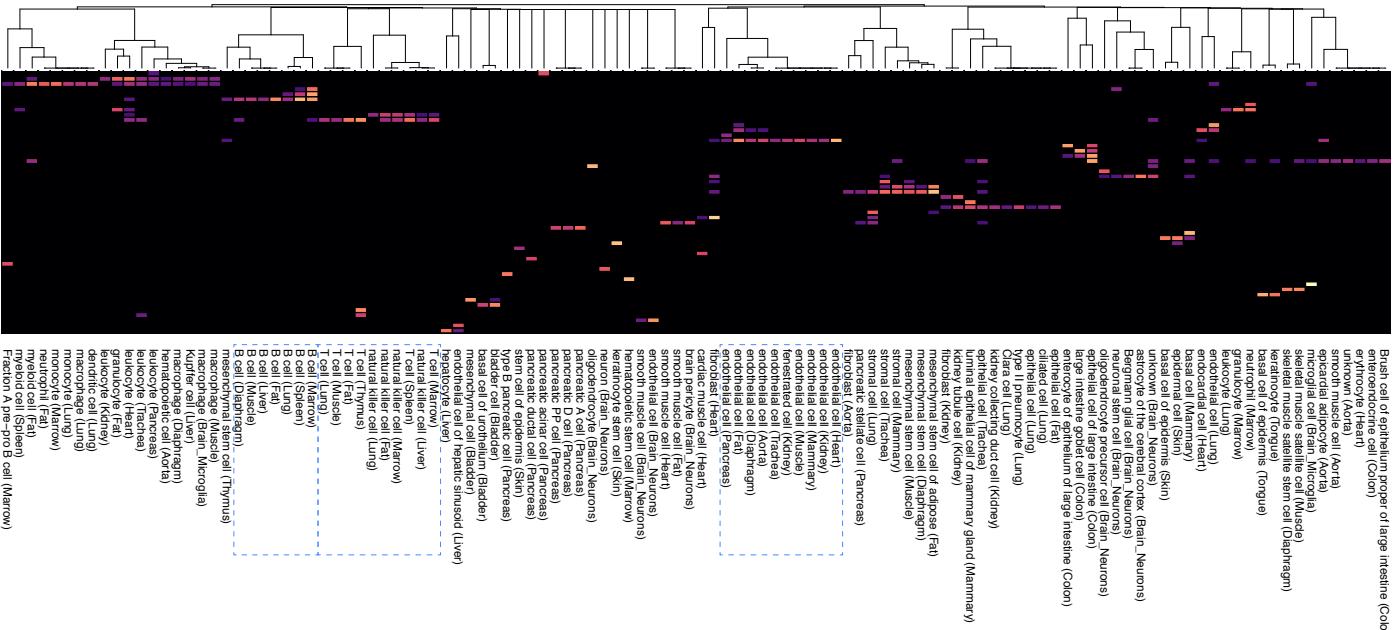
661 **Supplementary Figure 7** Cell type- and tissue-specific enrichment of gene expression
662 in TF gene clusters. Gene-gene correlations were calculated as in Figure 5 and clustered
663 by cutting a hierarchical clustering tree at a height of 1.5 (Euclidean distance of
664 correlations). To calculate scores, log10 cpm values were scaled to [0,1] for each gene.
665 The score on the x-axis is the log of the geometric mean of all of the TFs in the cluster
666 across all the cells of a cell type and tissue. The TFs in the cluster are indicated at the top
667 of the plot; for clusters with more than 5 genes, only the first 5 are shown. The
668 distribution of scores across all unique cell types and tissues is displayed, with the bar
669 plots colored by tissue origin of the cell type (colors as in Fig. 5). Gene clusters that are
670 highly specific show a strongly bimodal plot, with most cells expressing very low levels
671 of the gene and one or a small number expressing high levels (for example, the “Spdef
672 Atoh1 Creb3l4” cluster is highly specific to colon goblet cells). The names of the top
673 three cell types are displayed as a subtitle below the genes, with the highest-scored at the
674 top. Note that even though some gene clusters only separate one or two cell types, the top
675 3 by rank are still included in the subtitle.

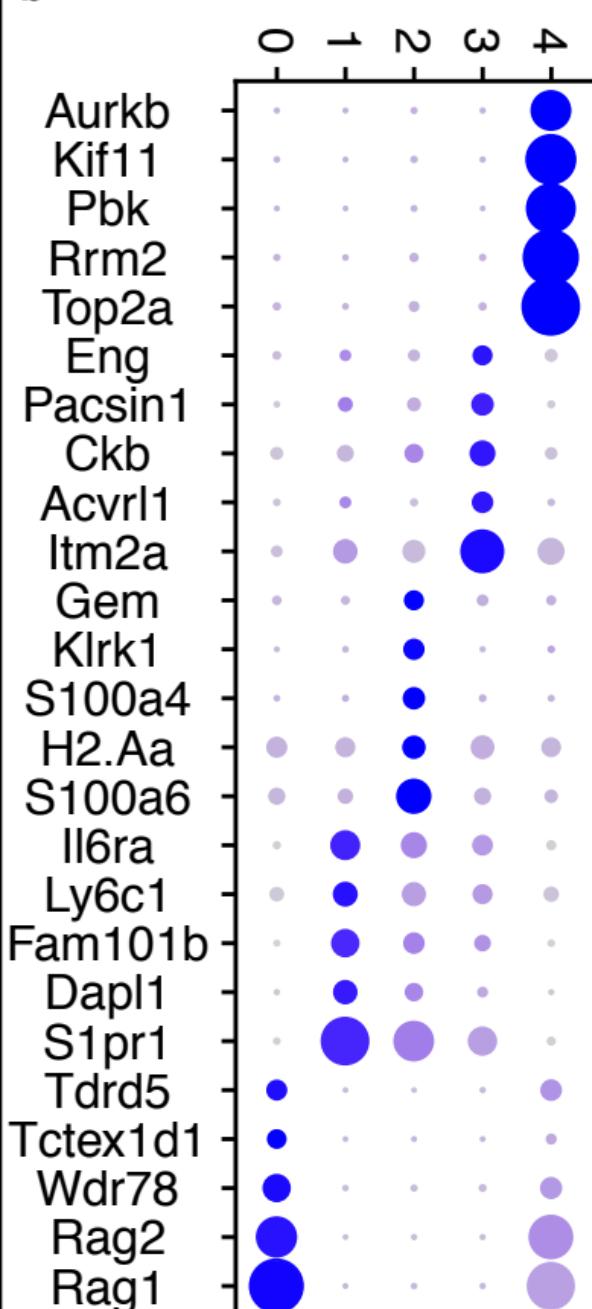
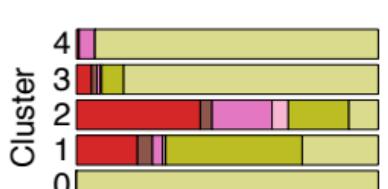
676

677

a**b****c****d****e****f****g**





a**b****c****d**