

27 **Summary**

28 Across the animal kingdom, gastrulation represents a key developmental event during
29 which embryonic pluripotent cells diversify into lineage-specific precursors that will
30 generate the adult organism. Here we report the transcriptional profiles of 116,312 single
31 cells from mouse embryos collected at nine sequential time-points ranging from 6.5 to 8.5
32 days post-fertilisation. We reconstruct a molecular map of cellular differentiation from
33 pluripotency towards all major embryonic lineages, and explore the complex events
34 involved in the convergence of visceral and primitive streak-derived endoderm.
35 Furthermore, we demonstrate how combining temporal and transcriptional information
36 illuminates gene function by single-cell profiling of *Tal1*^{-/-} chimeric embryos, with our
37 analysis revealing defects in early mesoderm diversification. Taken together, this
38 comprehensive delineation of mammalian cell differentiation trajectories *in vivo*
39 represents a baseline for understanding the effects of gene mutations during
40 development as well as a baseline for the optimisation of *in vitro* differentiation protocols
41 for regenerative medicine.

42

43 **Main Text**

44 The 48 hours of mouse embryonic development from embryonic day (E) 6.5 to E8.5
45 encompass the key phases of gastrulation and early organogenesis, when pluripotent
46 epiblast cells diversify into ectodermal, mesodermal and endodermal progenitors of all
47 major organs¹. Despite the central importance of this period of mammalian development,
48 we currently lack a comprehensive understanding of the underlying developmental
49 trajectories and molecular processes, principally because research efforts either
50 employed *in vitro* systems², focused on small numbers of genes, or limited the number of
51 developmental stages or cell types that were studied³.

52

53 ***A single-cell map of early embryogenesis***

54 To investigate the dynamic unfolding of cellular diversification during gastrulation and
55 early organogenesis, we complemented a previous E8.25 dataset⁵ by generating single-
56 cell RNA-seq (scRNA-seq) profiles from over 350 whole mouse embryos, collected at six-
57 hour intervals between E6.5 and E8.5 (Fig. 1a, b; Extended Data Fig. 1, 2a). Our dataset
58 thus captures Theiler stages TS9, TS10, TS11 and TS12, enriched in the Pre-Streak to
59 Early Streak, Mid-Streak to Late-streak, Neural Plate, and Headfold to Somitogenesis
60 stages, respectively⁶.

61

62 116,312 single-cell transcriptomes passed stringent quality control measures, with a
63 median of 3,436 genes detected per cell (Methods; Extended Data Fig. 2b-d;
64 Supplementary Information Table 1). Clustering and annotation identified 37 major cell

65 populations (Fig. 1c; Extended Data Fig. 2e), whose presence was coupled with
66 progression along the densely sampled time-points (Extended Data Fig. 3a-d). The
67 frequency of pluripotent epiblast cells declined over time, and mesodermal and definitive
68 endodermal lineages appeared as early as E6.75. Later, ectodermal lineages emerged
69 alongside a striking diversification of cell types from each germ layer at the onset of
70 organogenesis (Fig. 1d).

71
72 Transcriptional similarities between clusters (Methods; Extended Data Fig. 3e, f) were
73 consistent with prior knowledge: epiblast was similar to neuroectoderm and primitive
74 streak, with the latter being related to mesoderm and endoderm, consistent with the
75 divergence of the three germ layers. Neural and mesodermal layers were connected
76 during organogenesis (E8.25-E8.5) via a neuro-mesodermal progenitor population, which
77 has been reported to give rise to both caudal and neural tissues of the spinal cord
78 (Extended Data Fig. 3e)^{7,8}. Our atlas can be explored via an interactive and user-friendly
79 website: <https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018/>.

80

81 ***Mapping endoderm development***

82 Previous lineage tracing studies^{3,9} have shown that extra-embryonic and intra-embryonic
83 endodermal cells intercalate to form a single tissue, highlighting the plasticity of
84 embryonic cells (Extended Data Fig. 4a). Since extra-embryonic structures were sampled
85 alongside the gastrulating embryo, our dataset provided an opportunity to investigate this

86 convergence of primitive streak-derived definitive endoderm (DE) with visceral
87 endoderm-derived cells (VE) at the molecular level.

88

89 To this end, we performed a focused analysis using only the visceral endoderm, anterior
90 primitive streak, definitive endoderm and gut cell types (Fig. 2a; 5,015 cells), which was
91 consistent with gut endoderm arising from visceral as well as definitive endoderm,
92 identified by expression of *Ttr* and *Mixl1*, respectively (Extended Data Fig. 4b, c).
93 Inspection of the time-points of cell collection supported the transcriptional convergence
94 of these two lineages during development (Fig. 2b).

95

96 To define the transcriptional diversity within the mature gut, we exclusively analysed cells
97 collected at E8.25 and E8.5 and identified seven clusters corresponding to different cell
98 populations that line the gut tube (Fig. 2c), spanning the pharyngeal endoderm
99 (expressing *Nkx2-5*), foregut (expressing *Pyy*), midgut (expressing *Nepn*), and hindgut
100 (expressing *Cdx2*) (Extended Data Fig. 4d-g). Notably, foregut was split into two clusters,
101 named Foregut 1 and Foregut 2, likely corresponding to liver and lung precursors,
102 respectively (see hepatic-associated genes *Hhex*, *Sfrp5* and *Ttr*¹⁰⁻¹², and lung-associated
103 genes *Ripply3* and *Irx1*^{13,14}, in Fig. 2d). Hindgut cells were also divided into two distinct
104 clusters, Hindgut 1 and Hindgut 2, with significantly higher expression of the X-linked
105 genes *Trap1a* and *Rhox5* in Hindgut 1 (Fig. 2d, e). Given the spatial complexity of the gut
106 tube, we derived a pseudo-spatial ordering of these clusters using diffusion pseudotime
107 (DPT¹⁵), which recapitulated their anterior-posterior distribution (Fig. 2f).

108

109 To assess how VE cells may contribute to the mature gut, we inferred cellular transitions
110 along sequential collection time-points using transport maps¹⁶ (Methods). We then asked
111 whether cells in each gut cluster at E8.25 and E8.5 were likely derived from E7.0 VE or
112 DE ancestors. To account for cells with a permanent extra-embryonic fate, we added ExE
113 endoderm to the analysis (Methods; Extended Data Fig. 4h). Both the ExE endoderm and
114 the Hindgut 1 clusters were comprised of cells derived primarily from the VE, with all
115 remaining clusters having considerably smaller contributions (Fig. 2g; Extended Data Fig.
116 4i). Of note, the Hindgut 1 specific genes *Trap1a* and *Rhox5* are also expressed in ExE
117 endoderm and ExE ectoderm, consistent with the extra-embryonic origin of Hindgut 1
118 (Extended Data Fig. 4j). This suggests that, while Hindgut 1 and Hindgut 2 share a core
119 hindgut signature, Hindgut 1 cells also retain a transcriptional legacy from their extra-
120 embryonic origin. We also extended previous lineage tracing at E8.75⁹ by showing
121 enrichment of Ttr-YFP traced cells in the most posterior section of E8.5 embryos
122 containing the hindgut, using a *Ttr::Cre* transgene coupled with a conditional YFP
123 transgene in the ROSA26 locus (Extended Data Fig. 4k).

124

125 Next, we inferred which cells belonged to the developmental trajectories from the VE to
126 Hindgut 1, and from the DE to Hindgut 2 (Methods; Fig. 2h; Extended Data Fig. 5a, b),
127 ordered the cells using DPT¹⁵, and clustered genes based on their expression dynamics
128 along each trajectory (Methods; Supplementary Information Table 2; Extended Data Fig.
129 5c). We divided each trajectory into two domains: before and after completion of
130 endoderm intercalation at E7.5¹⁷. In the VE-Hindgut 1 trajectory, we observed

131 upregulation of VE genes during the first domain, followed by an abrupt decline as cells
132 proceeded towards the gut fate (Extended Data Fig. 5d), indicating that we captured a
133 subset of VE cells undergoing visceral maturation prior to the onset of DE intercalation.

134

135 Across both trajectories, a common set of genes were upregulated during intercalation
136 (Extended Data Fig. 5c, e), including genes involved in epithelial remodelling such as
137 *Pcna*, *Epcam*, and *Vim*, consistent with epithelial rearrangement⁹. Genes commonly
138 upregulated during the subsequent gut maturation and morphogenesis phase (Extended
139 Data Fig. 5c, f) were enriched for transcription factors (>20% of overlapping genes), 66%
140 of which were homeodomain proteins that showed sequential activation profiles,
141 indicative of a temporal collinearity during hindgut specification¹⁸. Analysis of dynamic
142 gene expression also revealed transcription factors specifically induced early in the VE-
143 Hindgut1 trajectory, including *Hes1*, *Pou5f1*, and *Sox4*, which represent promising
144 candidates for further study (Extended Data Fig. 5g).

145

146 ***Origins of haemato-endothelial lineages***

147 Red blood cells are formed in two consecutive waves in the yolk sac (YS), the first arising
148 at around E7.5 and the second from E8.25¹⁹. The first wave (primitive) generates
149 nucleated erythrocytes, which disappear shortly after birth. The second wave (YS
150 definitive) starts with the emergence of erythro-myeloid progenitors (EMPs) from YS
151 haemogenic endothelium (HE). These later migrate to the foetal liver and generate
152 definitive erythrocytes¹⁹ (Extended Data Fig. 6a).

153

154 While some key phenotypic and molecular distinctions between primitive and YS
155 definitive haematopoiesis are known, the respective *in vivo* progenitors are poorly
156 understood due to limiting cell numbers and lack of markers. To characterise these
157 processes more deeply, we computationally isolated and re-clustered cells assigned to
158 the erythroid, haemato-endothelial, blood progenitor, endothelial and mixed mesoderm
159 groups (15,875 cells; Fig. 3a, b; Extended Data Fig. 6b).

160

161 This analysis highlighted a putative trajectory towards the primitive erythroid lineage,
162 passing through clusters Haem1-2 (haemato-endothelial progenitors), BP1-2 (blood
163 progenitors) and Ery1-4 (erythrocytes) (Fig. 3a, b). This trajectory did not include the
164 endothelial region (clusters EC1-8), which was enriched for cells collected at E8.25-E8.5,
165 displayed a complex structure, and expressed high levels of *Kdr*, which encodes the
166 protein FLK1 (termed *Kdr*^{hi} region hereafter) (Fig. 3c). Notably, some of these endothelial
167 cells expressed haematopoietic markers, such as *Spi1 (Pu.1)* and *Itga2b* (Fig. 3d),
168 potentially highlighting the emergence of blood from endothelium during the second
169 wave²⁰. Incorporating temporal information (Fig. 3b) suggested that, unlike the second
170 haematopoietic wave, the first wave does not transit through a molecular state with
171 classical mature endothelial characteristics, as marked by *Cdh5* and *Pecam1* (Extended
172 Data Fig. 6c).

173

174 Endothelial cells are generated independently in the YS, allantois and embryo proper
175 (EP)^{21,22}, with the allantoic endothelium being hypothesised to display a specific
176 transcriptional signature⁵. To test whether the heterogeneity in the *Kdr*^{hi} region was
177 associated with different anatomical locations, we dissected out the YS, allantois and EP
178 from a new batch of E8.25 embryos, purified endothelial cells by flow sorting the FLK1⁺
179 population, and performed scRNA-seq on 288 cells using Smart-seq2²³ (Extended Data
180 Fig. 6d-f). Assigning the cells from the *Kdr*^{hi} region (EC1-8) to their most likely embryonic
181 location intimated that diverse anatomical origin can partially explain the transcriptional
182 heterogeneity observed in the endothelium (Fig. 3e).

183

184 Previous *in vitro* colony forming assays of early embryonic cells suggested that, in
185 addition to erythrocytes, the primitive wave also gives rise to macrophage and
186 megakaryocytic progenitors²⁴⁻²⁶. However, the molecular nature of these progenitors
187 remains obscure. In our atlas, we identified two rare cell groups (present at a frequency
188 of around 0.1%) that we annotated as megakaryocytes (Mk) and myeloid cells (My;
189 Extended Data Fig. 6g), providing an opportunity to characterise their molecular profiles
190 based on primary *in vivo* cells (Fig. 3f).

191

192 Recent reports suggest that early myeloid progenitors can give rise to brain microglia²⁷.
193 Consistent with this, cells in our My population expressed *Ptprc* (CD45), *Kit*, *Csf1r* and
194 *Fcgr3* (CD16/32), previously reported as markers of the E8.5 EMP-like population that
195 give rise to microglial macrophages^{20,28} (Fig. 3g). However, they did not express

196 appreciable levels of more mature microglial-related genes such as *Cx3cr1*, *Adgre1*
197 (F4/80) and *Tmem119*^{29,30}. To investigate the location and frequency of these cells in the
198 embryo, we dissected different regions of E8.5 embryos (Extended Data Fig. 6h) and
199 performed flow cytometry analysis using the markers CD16/32 and CSF1R. Rare
200 CD16/32⁺CSF1R⁺ cells were found in all dissected regions (Extended Data Fig. 6i),
201 indicating that by E8.5 this population has already started to migrate out of the YS.

202

203 ***A platform to dissect genetic mutations***

204 Previous work has emphasised the critical role of the bHLH transcription factor TAL1
205 (SCL) in haematopoiesis, with *Tal1*^{-/-} mouse embryos dying around E9.5 from severe
206 anaemia³¹. Dissecting the temporal and mechanistic roles *in vivo* of major regulatory
207 genes is challenging using knockout mice, as they require time-consuming breeding and
208 genotyping of embryos. Additionally, direct effects of the mutation are often masked by
209 gross developmental malformation or embryo lethality. To circumvent these difficulties,
210 we generated chimeric mouse embryos where *Tal1*^{-/-} tdTomato⁺ mouse embryonic stem
211 cells (ESC) were injected into wildtype blastocysts. In the resulting chimeras, wildtype
212 cells still produce blood cells, allowing the specific effects of TAL1 depletion to be studied
213 in an otherwise healthy embryo³².

214

215 To determine whether *Tal1* mutant cells were associated with abnormalities in specific
216 lineages, we sorted tdTomato⁻ (WT) and tdTomato⁺ (*Tal1*^{-/-}) cells from chimeric embryos
217 at E8.5 followed by scRNA-seq (Fig. 4a; Extended Data Fig. 7a, b). Each cell was

218 annotated by computationally mapping its transcriptome onto our wildtype atlas (Methods;
219 Fig. 4b; Extended Data Fig. 7c-e). Consistent with the pivotal role of *Tal1* in
220 haematopoiesis, tdTomato⁺ cells did not contribute to blood lineages (Fig. 4b; Extended
221 Data Fig. 7e-g). Importantly, we confirmed that wildtype control tdTomato⁺ *Tal1*^{+/+} ESCs,
222 when injected into wildtype embryos, make a similar contribution to haematopoiesis as
223 the tdTomato⁻ host cells (Extended Data Fig. 7h, i).

224

225 Comparisons between WT and *Tal1*^{-/-} chimeric cells mapped to the landscape defined in
226 Fig. 3a illustrated that TAL1 depletion disrupts the emergence of primitive erythroid cells
227 as well as our newly characterised megakaryocyte and myeloid cells (Fig. 4c). Although
228 a subset of *Tal1*^{-/-} cells were mapped to the haemogenic endothelial groups EC6 and
229 EC7, they did not express genes associated with blood development, such as *Itga2b* or
230 the known TAL1 target gene *Cbfa2t3* (*Eto2*), in contrast to the host WT cells, supporting
231 the disruption of the second haematopoietic wave upon TAL1 depletion (Fig. 4d).

232

233 To further characterise the developmental block in the second haematopoietic wave, we
234 quantified the relative contributions of *Tal1*^{-/-} and WT chimeric cells to each “EC” and
235 “Haem” cluster described in Fig. 3 (Fig. 4e; Supplementary Information Table 3).
236 Interestingly, E8.5 *Tal1*^{-/-} cells were more abundant than the WT cells in EC3, one of the
237 earliest-appearing endothelial sub-clusters (Fig. 3b). While mutant cells might simply
238 accumulate in this state, *Tal1*^{-/-} cells mapped to EC3 may alternatively acquire a novel
239 transcriptional state, which is similar but not identical to EC3. To clarify this, we performed

240 differential expression analyses comparing EC3-mapped *Tal1*^{-/-} cells to their most similar
241 cells in the reference atlas, and to the WT host chimeric cells mapped to EC3.
242 Interestingly, we observed a small set of genes specifically upregulated in EC3-mapped
243 *Tal1*^{-/-} cells, including *Pcolce*, *Tdo2* and *Plagl1* (Fig. 4f; Extended Data Fig. 8a). When
244 inspecting expression of these genes in our atlas, we observed high expression in the
245 mesenchyme and other mesoderm clusters, such as the allantois, paraxial, pharyngeal
246 and intermediate mesoderm (Extended Data Fig. 8b). Moreover, we noted that a subset
247 of these cells also upregulated cardiac-related genes, such as *Nkx2-5*, *Mef2c* and *Tnnt2*
248 (Fig. 4f), consistent with a previous report that *Tal1*^{-/-} YS cells can adopt a cardiomyocyte-
249 like phenotype³³. However, these cells did not present a full cardiomyocyte transcriptional
250 program and continued to express endothelial genes such as *Esam* and *Sox17*, albeit
251 with some down-regulation compared to their EC3 atlas counterparts. These results
252 suggest that *Tal1* disruption blocks cells at a transcriptional state similar to that of the
253 EC3 cluster during the second wave of blood development. Moreover, when unable to
254 proceed towards a haemogenic phenotype, EC3-mapped *Tal1*^{-/-} cells begin to activate
255 other mesodermal programs. This is consistent with prior evidence that haematopoietic
256 precursors isolated from E7.5 mouse embryos are endowed with mesodermal plasticity
257 when cultured *ex vivo*³⁴.

258

259 **Discussion**

260 Our comprehensive atlas of mouse gastrulation and early organogenesis offers a
261 powerful resource for investigating the molecular underpinnings of cell fate decisions
262 during this key period of mammalian development. We exploited this resource by

263 investigating two specific developmental phenomena: the transdifferentiation process of
264 visceral endoderm cells contributing to the composition of the embryonic gut, and the
265 emergence of rare blood cells in the early embryo. Moreover, we used our atlas as a
266 reference for the analysis of *Tal1*^{-/-} mutant chimeric embryos, which highlighted where
267 TAL1 is critical for progression into the blood lineage, and also identified a novel
268 transcriptional state unique to *Tal1*^{-/-} cells, where genes from multiple different
269 mesodermal tissues are expressed alongside endothelial genes.

270

271 More broadly, our chimera analysis illustrates the utility and efficiency of such a model for
272 studying the molecular and cellular consequences of a wide range of developmental
273 mutants, including those that are embryonically lethal and relevant for human
274 developmental disorders. In sum, our work, in a widely-used and experimentally relevant
275 mammalian system, complements recent single-cell expression profiling surveys in early
276 zebrafish and *Xenopus* embryos³⁵⁻³⁷. Collectively, these studies demonstrate that densely
277 sampled large-scale single-cell profiling has the potential to advance our understanding
278 of embryonic development in vertebrates.

279

280 **Acknowledgements**

281 We thank William Mansfield for blastocyst injections, Aaron T.L. Lun and Fiona Hamey
282 for discussions concerning the analysis, Tina L. Hamilton for technical support in embryo
283 collection, Sarah Kinston and Kenneth Jones for technical assistance, the Flow Cytometry
284 Core Facility at CIMR for cell sorting, the CRUK-CI genomics core for the chimera scRNA-

285 seq 10X libraries and for letting us use the 10X Chromium after hours, and the Wellcome
286 Sanger Institute DNA Pipelines Operations for sequencing. We thank Kat Hadjantonakis
287 for generously sharing the *Ttr::Cre* mouse line. Research in the authors' laboratories is
288 supported by the Wellcome, MRC, CRUK, Bloodwise, NIH-NIDDK; by core support grants
289 from the Wellcome to the Cambridge Institute for Medical Research and Wellcome-MRC
290 Cambridge Stem Cell Institute; and by core funding from Cancer Research UK and the
291 European Molecular Biology Laboratory. B.P.-S. and D.L.L.H. are funded by the
292 Wellcome Trust 4-Year PhD Programme in Stem Cell Biology and Medicine and the
293 University of Cambridge, UK. D.L.L.H is also funded by the Cambridge Commonwealth
294 European and International Trust. J.A.G. is funded by the Wellcome Trust Mathematical
295 Genomics and Medicine Programme at the University of Cambridge [109081/Z/15/A].
296 C.G. is funded by the Swedish Research Council [2017-06278]. This work was funded as
297 part of a Wellcome Strategic Award [105031/Z/14/Z] awarded to Wolf Reik, Berthold
298 Göttgens, John Marioni, Jennifer Nichols, Ludovic Vallier, Shankar Srinivas, Benjamin
299 Simons, Sarah Teichmann, and Thierry Voet, by Wellcome Trust grant [108438/Z/15]
300 awarded to John Marioni and by a BBSRC grant [BBS/E/B/000C0421] awarded to Wolf
301 Reik.

302

303 **Author contributions**

304 B.P.-S., W.J., F.J.C.-N., C.M., J.N. generated the atlas dataset. C.G. designed and
305 executed the chimera dataset generation and associated experiments. D.L.L.H. assisted
306 in the generation of the *Tal1^{-/-}* ESC line. J.A.G. performed pre-processing, low-level
307 analyses, batch-correction, clustering, and global visualisation of the atlas and chimera

308 datasets, and designed the associated website. B.P.-S. curated the clustering and
309 evaluated the connectivity between cell types. B.P.-S., C.G. annotated atlas cell types.
310 J.A.G., C.G. analysed atlas endoderm. B.P.-S assisted in the endoderm analyses by
311 generating force-directed layouts and inferring trajectories using *graph abstraction* as an
312 alternative approach. R.C.V.T. performed *Ttr::Cre* embryo imaging experiments. B.P.-S
313 analysed atlas haemato-endothelium, and performed associated experiments and
314 analyses. J.A.G. mapped chimera cells to the atlas. B.P.-S., C.G. analysed effects of *Tal1*^{-/-}.
315 T.W.H. contributed to the mapping and analysis of chimeras. X.I.-S. provided advice on
316 bioinformatics analysis. W.R., S.S., B.D.S., J.N., J.C.M., B.G. supervised the study. B.P.-
317 S., J.A.G., C.G., T.W.H., J.C.M., B.G. wrote the manuscript. All authors read and
318 approved the final manuscript.

319

320 **Author Information Statement**

321 Reprints and permissions information is available at www.nature.com/reprints. The
322 authors declare no competing interests. Correspondence and requests for materials
323 should be addressed to bg200@cam.ac.uk or john.marioni@cruk.cam.ac.uk.

324

325 **References**

- 326 1 Tam, P. P. L. & Behringer, R. R. Mouse gastrulation: the formation of a mammalian
327 body plan. *Mech. Dev.* **68**, 3-25 (1997).
- 328 2 Loh, K. M. *et al.* Mapping the Pairwise Choices Leading from Pluripotency to
329 Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451-467 (2016).

330 3 Viotti, M., Nowotschin, S. & Hadjantonakis, A.-K. SOX17 links gut endoderm
331 morphogenesis and germ layer segregation. *Nat. Cell Biol.* **16**, 1146-1156 (2014).

332 4 Lescroart, F. *et al.* Defining the earliest step of cardiovascular lineage segregation
333 by single-cell RNA-seq. *Science* **359**, 1177-1181 (2018).

334 5 Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution
335 reveals a role for the leukotriene pathway in regulating blood progenitor formation.
336 *Nat. Cell Biol.* **20**, 127 (2018).

337 6 Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by
338 morphological landmarks in the dissecting microscope. *Development* **118**, 1255-
339 1266 (1993).

340 7 Koch, F. *et al.* Antagonistic Activities of Sox2 and Brachyury Control the Fate
341 Choice of Neuro-Mesodermal Progenitors. *Dev. Cell* **42**, 514-526.e517 (2017).

342 8 Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J.-F.
343 Redefining the Progression of Lineage Segregations during Mammalian
344 Embryogenesis by Clonal Analysis. *Dev. Cell* **17**, 365-376 (2009).

345 9 Kwon, G. S., Viotti, M. & Hadjantonakis, A.-K. The Endoderm of the Mouse Embryo
346 Arises by Dynamic Widespread Intercalation of Embryonic and Extraembryonic
347 Lineages. *Dev. Cell* **15**, 509-520 (2008).

348 10 Finley, K. R., Tennessen, J. & Shawlot, W. The mouse secreted frizzled-related
349 protein 5 gene is expressed in the anterior visceral endoderm and foregut
350 endoderm during early post-implantation development. *Gene Expr Patterns* **3**,
351 681-684 (2003).

- 352 11 Makover, A., Soprano, D. R., Wyatt, M. L. & Goodman, D. S. An in situ-
353 hybridization study of the localization of retinol-binding protein and transthyretin
354 messenger RNAs during fetal development in the rat. *Differentiation* **40**, 17-25
355 (1989).
- 356 12 Martinez Barbera, J. P. *et al.* The homeobox gene Hex is required in definitive
357 endodermal tissues for normal forebrain, liver and thyroid formation. *Development*
358 **127**, 2433-2445 (2000).
- 359 13 Bosse, A. *et al.* Identification of the vertebrate Iroquois homeobox gene family with
360 overlapping expression during early development of the nervous system. *Mech.*
361 *Dev.* **69**, 169-181 (1997).
- 362 14 Osipovich, A. B. *et al.* Insm1 promotes endocrine cell differentiation by modulating
363 the expression of a network of genes that includes Neurog3 and Ripply3.
364 *Development* **141**, 2939-2949, doi:10.1242/dev.104810 (2014).
- 365 15 Haghverdi, L., Buttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion
366 pseudotime robustly reconstructs lineage branching. *Nat Methods* **13**, 845-848,
367 doi:10.1038/nmeth.3971 (2016).
- 368 16 Schiebinger, G. *et al.* Reconstruction of developmental landscapes by optimal-
369 transport analysis of single-cell gene expression sheds light on cellular
370 reprogramming. *bioRxiv*, 191056 (2017).
- 371 17 Viotti, M., Foley, A. C. & Hadjantonakis, A. K. Gutsy moves in mice: cellular and
372 molecular dynamics of endoderm morphogenesis. *Philos. Trans. R. Soc. Lond. B*
373 *Biol. Sci.* **369**, doi:10.1098/rstb.2013.0547 (2014).

374 18 Deschamps, J. & Duboule, D. Embryonic timing, axial stem cells, chromatin
375 dynamics, and the Hox clock. *Genes & Development* **31**, 1406-1416 (2017).

376 19 Palis, J. Hematopoietic stem cell-independent hematopoiesis: emergence of
377 erythroid, megakaryocyte, and myeloid potential in the mammalian embryo. *FEBS*
378 *Lett.* **590**, 3965-3974 (2016).

379 20 McGrath, K. E. *et al.* Distinct Sources of Hematopoietic Progenitors Emerge before
380 HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Reports*
381 **11**, 1892-1904 (2015).

382 21 Downs, K. M., Gifford, S., Blahnik, M. & Gardner, R. L. Vascularization in the
383 murine allantois occurs by vasculogenesis without accompanying erythropoiesis.
384 *Development* **125**, 4507-4520 (1998).

385 22 Patan, S. in *Angiogenesis in Brain Tumors Cancer Treatment and Research* 3-
386 32 (Springer, Boston, MA, 2004).

387 23 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single
388 cells. *Nature Methods* **10**, 1096-1098 (2013).

389 24 Palis, J., Robertson, S., Kennedy, M., Wall, C. & Keller, G. Development of
390 erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse.
391 *Development* **126**, 5073-5084 (1999).

392 25 Tober, J. *et al.* The megakaryocyte lineage originates from hemangioblast
393 precursors and is an integral component both of primitive and of definitive
394 hematopoiesis. *Blood* **109**, 1433-1441, doi:10.1182/blood-2006-06-031898
395 (2007).

- 396 26 Xu, M.-j. *et al.* Evidence for the presence of murine primitive megakaryocytopoiesis
397 in the early yolk sac. *Blood* **97**, 2016-2022 (2001).
- 398 27 Hoeffel, G. *et al.* C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes
399 Give Rise to Adult Tissue-Resident Macrophages. *Immunity* **42**, 665-678 (2015).
- 400 28 Perdiguero, E. G. *et al.* The Origin of Tissue-Resident Macrophages: When an
401 Erythro-myeloid Progenitor Is an Erythro-myeloid Progenitor. *Immunity* **43**, 1023-
402 1024, doi:10.1016/j.immuni.2015.11.022 (2015).
- 403 29 Bennett, M. L. *et al.* New tools for studying microglia in the mouse and human
404 CNS. *Proceedings of the National Academy of Sciences* **113**, E1738-E1746
405 (2016).
- 406 30 Ginhoux, F. *et al.* Fate Mapping Analysis Reveals That Adult Microglia Derive from
407 Primitive Macrophages. *Science* **330**, 841-845 (2010).
- 408 31 Shivdasani, R. A., Mayer, E. L. & Orkin, S. H. Absence of blood formation in mice
409 lacking the T-cell leukaemia oncoprotein tal-1/SCL. *Nature* **373**, 432-434 (1995).
- 410 32 Robb, L. *et al.* The scl gene product is required for the generation of all
411 hematopoietic lineages in the adult mouse. *The EMBO journal* **15**, 4123-4129
412 (1996).
- 413 33 Van Handel, B. *et al.* Scl represses cardiomyogenesis in prospective hemogenic
414 endothelium and endocardium. *Cell* **150**, 590-605, doi:10.1016/j.cell.2012.06.026
415 (2012).
- 416 34 Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. Haemangioblast
417 commitment is initiated in the primitive streak of the mouse embryo. *Nature* **432**,
418 625-630 (2004).

419 35 Briggs, J. A. *et al.* The dynamics of gene expression in vertebrate embryogenesis
420 at single-cell resolution. *Science*, eaar5780 (2018).

421 36 Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during
422 zebrafish embryogenesis. *Science*, eaar3131 (2018).

423 37 Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and
424 lineage in the zebrafish embryo. *Science*, eaar4362 (2018).

425 38 Pijuan-Sala, B., Guibentif, C. & Göttgens, B. Single-cell transcriptional profiling: a
426 window into embryonic cell-type specification. *Nature Reviews Molecular Cell*
427 *Biology*, 1 (2018).

428

429 **Main Figure legends**

430 **Figure 1: A single-cell resolution atlas of mouse gastrulation and early**
431 **organogenesis. a**, Overview of embryonic developmental time-points sampled,
432 alongside corresponding Theiler and Downs and Davies stages. Adapted from³⁸.
433 Numbers indicate days post-fertilisation. PrS: Pre-Streak, ES: Early Streak, MS: Mid-
434 Streak, LS: Late Streak, OB: Neural Plate no bud, EB: Neural Plate Early Bud, LB: Neural
435 Plate Late Bud, EHF: Early Headfold, LHF: Late Headfold, 1-7S: 1-7 Somites. **b**,
436 Representative images of sampled embryos (see Supplementary Information Table 1 for
437 sample collection and size). Scale bars: 0.25 mm. **c**, UMAP plot showing all cells of the
438 atlas (116,312 cells). Cells are coloured by their cell type annotation and numbered
439 according to the legend below. ExE: Extra-embryonic, NMP: Neuromesodermal
440 progenitors, PGC: Primordial germ cells, prog.: progenitor, Def.: Definitive. **d**, Change in

441 frequency of cell type per time-point, displaying a progressive increase in cell type
442 complexity throughout our sampling.

443

444 **Figure 2: Molecular conversion and subsequent diversification during early**
445 **endoderm development.** **a-c**, Force-directed graph layout of the endoderm cell subset
446 (5,015 cells) coloured by (a) global cell type annotation, (b) embryo collection time-point,
447 (c) mature gut cell types. Each point is a cell, and cells close to each other have similar
448 transcriptional profiles. **d**, Heatmap illustrating mean expression of marker genes for each
449 mature gut cluster (row-normalised). **e**, Volcano plot showing differentially-expressed
450 genes between Hindgut 1 (53 cells) and 2 (148 cells). Red: significantly differentially-
451 expressed genes (BH-adjusted $p < 0.1$; Methods). The five most significantly differentially-
452 expressed genes in each direction are labelled. **f**, Pseudo-spatial ordering of cells along
453 the gut tube. x-axis: pseudo-space coordinate corresponding to DPT values. **g**, Fraction
454 of cells of each mature gut cluster predicted to derive from visceral endoderm (VE) or
455 definitive endoderm (DE). **h**, Force-directed graph coloured by putative trajectories for
456 formation of the Hindgut clusters.

457

458 **Figure 3: Temporal analysis of blood emergence reveals early myeloid cells.** **a**,
459 Force-directed graph layout of cells associated with the blood lineage, coloured by
460 subcluster (15,875 cells). Box shows a zoomed section focusing on Myeloid (My),
461 Megakaryocytic (Mk), and haemogenic endothelial cells. **b**, *Graph abstraction*
462 summarising the relationships between the subclusters as in **(a)**, coloured by subcluster

463 (left) and collection time-point (right), excluding two samples of mixed-time-point
464 embryos. **c**, Expression levels of *Kdr*, overlaid on the force-directed layout. **d**, Expression
465 levels of *Spi1* and *Itga2b*, overlaid on the inset of the force-directed layout. **e**, Fraction of
466 EC cells mapped to yolk sac, allantois and embryo proper. **f**, Heatmap illustrating row-
467 normalised expression of genes upregulated in EC7 (197 cells), Haem4 (102 cells), My
468 (56 cells), BP4 (54 cells) and Mk (32 cells) clusters when comparing all subclusters in **(a)**
469 ($\log_{2}FC > 2.5$; BH-adjusted $p < 0.05$). **g**, Heatmap illustrating the log-count expression
470 (ranging from 0 to 3.5) of previously described microglial (Micr.) and EMP markers.

471

472 **Figure 4: Mapping *Tal1*^{-/-} chimeras to the atlas identifies molecular states**
473 **associated with defects in haemato-endothelial development.** **a**, Experimental
474 design for *Tal1*^{-/-} chimera generation and sequencing. **b**, UMAPs of chimera cells (25,078
475 WT cells; 26,326 *Tal1*^{-/-} cells). Points are coloured and numbered according to their
476 computationally assigned cell type, as in Figure 1. Numbers and legend have only been
477 specified for those cell types that are clearly visible in the plot. White arrowhead highlights
478 blood cells, which are depleted in mutant cells. **c**, Mapping of blood-related cells from the
479 chimera onto the blood-related cells from the atlas. Left: WT (9,336 cells); right: *Tal1*^{-/-}
480 (2,911 cells). Arrowheads denote the position at which blood development appears
481 blocked in *Tal1*^{-/-} cells. **d**, Heatmap illustrating the row-normalised expression of blood
482 (*Cbfa2t3* and *Itga2b*) and endothelial (*Cdh5* and *Pecam1*) genes in EC6 WT (43 cells),
483 EC6 *Tal1*^{-/-} (28 cells), EC7 WT (117 cells), and EC7 *Tal1*^{-/-} (7 cells) cells. **e**, Log-fold-
484 change abundance of *Tal1*^{-/-} cells with respect to WT chimera cells in each of the clusters.
485 Below are the absolute numbers of cells resulting from the injected *Tal1*^{-/-} cells (red) and

486 from the host WT (black). **f**, Heatmap illustrating the row-normalised expression of genes
487 upregulated in EC3-mapped *Tal1*^{-/-} cells. From left to right, columns represent a sample
488 of atlas cardiomyocytes (CM) (200 cells), *Tal1*^{-/-} EC3 (328 cells), (3) WT EC3 (23 cells)
489 and (4) atlas EC3 cells (107 cells). Illustrative genes have been manually selected from
490 the full heatmap shown in Extended Data Fig. 8a.

491

492 **Methods**

493 **Embryo collection and sequencing.**

494 All procedures were performed in strict accordance to the UK Home Office regulations for
495 animal research. Chimeric mouse embryos were generated under the project licence
496 number PPL 70/8406.

497 **Reference atlas.** Pregnant C57BL/6 mouse females were purchased from Charles River
498 and delivered one day before or on the day of embryo harvest. Mouse embryos were
499 dissected at time-points E6.5, E6.75, E7.0, E7.25, E7.5, E7.75, E8.0, E8.25 and E8.5. As
500 previously reported⁶, development can proceed at different speeds between embryos,
501 even within the same litter (Fig. 1a; Extended Data Fig. 1). Consequently, we adopted
502 careful staging by morphology (Downs and Davies staging⁶) to exclude clear outliers.
503 Following euthanasia of the females using cervical dislocation, the uteri were collected
504 into PBS with 2% heat-inactivated FCS and the embryos were immediately dissected and
505 processed for scRNA-seq. Two samples contained pooled embryos staged across
506 several time-points. Cells from these samples are denoted as “Mixed” in Figures, and
507 “mixed_gastrulation” in Supplementary Information Table 4. Embryos from the same
508 stage were pooled to make individual 10X samples, and single-cell suspensions were
509 prepared by incubating the embryos with TrypLE Express dissociation reagent (Life
510 Technologies) at 37 °C for 7 min and quenching with heat inactivated serum. The resulting
511 single-cell suspension was washed and resuspended in PBS with 0.4% BSA, and filtered
512 through a Flowmi Tip Strainer with 40 µm porosity (ThermoFisher Scientific,
513 #136800040). Cell counts were then assessed with a haemocytometer. scRNA-seq
514 libraries were subsequently generated using the 10X Genomics Chromium system

515 (version 1 chemistry) and samples were sequenced according to the manufacturer's
516 instructions on an Illumina HiSeq 2500 platform. Supplementary Information Table 1
517 contains detailed information on embryo collection, and Supplementary Information Table
518 4 contains metadata for each sequenced cell. Sample sizes were chosen to maximise the
519 number of recovered cells from each experiment and to obtain total cell numbers similar
520 to the estimated cell numbers in mouse embryos at their respective stages. The sample
521 sizes were also dependent on the number of viable embryos from each litter. Cells were
522 partitioned to prevent overloading of a single 10X lane.

523 **Yolk sac, allantois and embryo proper EC dissection experiment.** Mice were bred
524 and maintained at the University of Cambridge, in individually ventilated cages with sterile
525 bedding; sterile food and water were provided *ad libitum*. All animals were kept in
526 pathogen-free conditions. Timed-matings were set up between C57BL/6 mice, purchased
527 from Charles River. Upon dissection, only embryos staged as Theiler Stage 12 were
528 further processed. Allantois, yolk sac and embryo proper were dissected and placed into
529 separate tubes. Single-cell suspensions were prepared by incubating the embryos with
530 TrypLE Express dissociation reagent (Life Technologies) at 37°C for 7 min and quenching
531 with heat inactivated serum. Single cells were subsequently stained with FLK1-PE
532 antibody (1:100; Biolegend, cat# 12-5821-83, clone Avas12a1, lot# E01819-1631) and
533 DAPI as viability stain (1 µg/ml; Sigma). Live FLK1⁺ cells were isolated by fluorescence-
534 activated cell sorting (FACS) using a BD Influx sorter into individual wells of a 96-well
535 plate containing lysis buffer (0.2% (v/v) Triton X-100 and 2 U/µl SUPERase-In (Invitrogen,
536 #AM2696) and stored at -80 °C (1 plate per tissue was prepared). Plates were processed
537 following the Smart-seq2 protocol as previously described²³ and libraries were generated

538 using the Illumina Nextera XT DNA preparation kit. Libraries were pooled and sequenced
539 on an Illumina HiSeq 4000. Sample sizes were chosen based on the amount of viable
540 endothelial cells recovered from the experiment and we aimed to have an equal (or very
541 similar) number of endothelial cells from each of the dissected regions that was large
542 enough (i.e. 96 per sample) to infer correlations with the atlas dataset.

543 **Flow cytometry analysis of myeloid progenitors.** Mice and embryos were obtained as
544 above (*Yolk sac, allantois and embryo proper EC dissection experiment*). Yolk sac,
545 allantois, amnion, head, heart and trunk were dissected and placed into separate tubes.
546 Single-cell suspensions were prepared as above (*Yolk sac, allantois and embryo proper*
547 *EC dissection experiment*). Single cells were subsequently stained with CD16/32-BV711
548 (1:200; Biolegend, cat# 101337, clone 93, lot# B251800) for 20 min at 4°C, washed with
549 2 ml PBS+2%FCS, blocked with Fc block CD16/32 (1:100; eBioscience, cat# 14-0161-
550 85, clone 93, lot# E03558-1640) and stained with CSF1R-BV605 (1:800; Biolegend; cat#
551 135517, clone AFS98, lot# B196541) for 30 minutes at 4°C. Cells were then washed and
552 7AAD was added as a viability stain (1:200; BD Pharmigen; cat# 51-68981E, lot#
553 7061885). Cells were analysed using a BD Fortessa cytometer. Gates were established
554 using Fluorescence Minus One (FMO) controls. Two biological replicates were
555 performed: one pool of 12 and one pool of 13 embryos.

556 **Ttr-YFP embryo staining.** *Ttr::Cre* stud male mice⁹ were crossed with R26R-YFP
557 females³⁹. Dissected E8.5 embryos were fixed for 1 hour at room temperature with 4%
558 paraformaldehyde (PFA) in Phosphate buffered saline (PBS). The embryos were then
559 washed 3 times in PBS with 0.1% Triton X-100 (PBT-0.1%) for 15 minutes, permeabilised
560 in PBT-0.25% for 40 minutes and washed again 3 times in PBT-0.1%. The embryos were

561 transferred to blocking solution (5% donkey serum (Sigma, #D9663), 1% Bovine Serum
562 Albumin (Sigma, #A7906) in PBT-0.1%) overnight (o/n) at 4°C. Primary antibody (Chicken
563 Anti-GFP; 1:100; Abcam, cat# ab13970, Lot# GR3190550-2) was then added in blocking
564 solution and incubated o/n at 4°C. The embryos were washed 3 times in PBT-0.1% and
565 incubated o/n at 4°C in PBT-0.1% with the secondary antibody (Goat Anti-chicken 488;
566 Sigma; 1:100; cat# A11039; Lot# 1899514) and Phalloidin 555 (1:100; Sigma; #19083),
567 then subsequently washed 3 times PBT-0.1% for 15 min and mounted in Vectashield
568 mounting media with DAPI for at least 24 hrs at 4°C. Images were captured using a Zeiss
569 880 confocal microscope.

570

571 **Chimera generation and sequencing.**

572 TdTomato-expressing mouse embryonic stem cells (ESC) were derived as previously
573 described⁴⁰ from E3.5 blastocysts obtained by crossing a male ROSA26tdTomato (Jax
574 Labs – 007905) with a wildtype C57BL/6 female. The cells were negative for mycoplasma
575 contamination. The cells were expanded under the 2i+LIF conditions⁴¹ and transiently
576 transfected with a Cre-IRES-GFP plasmid⁴² using Lipofectamine 3000 Transfection
577 Reagent (ThermoFisher Scientific, #L3000008) according to manufacturer's instructions.
578 Single GFP⁺ cells were sorted 48h post-transfection into 96-well plates. Individual clones
579 were allowed to grow and were manually picked for expansion. A tdTomato-positive,
580 male, karyotypically normal line, competent for chimera generation as assessed using
581 morula aggregation assay, was selected for targeting *Ta11*. Two guides targeting exon 4
582 were designed using the <http://crispr.mit.edu> tool (guide 1:
583 GAACCCACTATGGAAAGAGA; guide 2: GAGGCCCTCCCATATGAGA) and were

584 cloned into the pX458 plasmid (Addgene, #48138) as previously described⁴³. The
585 resulting plasmids were then used to transfect the cells as detailed above. Single
586 transfected clones were expanded and assessed for Cas9-induced mutations. Genomic
587 DNA was isolated by incubating cell pellets in 0.1 mg/ml of Proteinase K (Sigma,
588 #03115828001) in TE buffer at 50°C for 2 hrs, followed by 5 min at 99°C. The sequence
589 flanking the guide-targeted sites was amplified from the genomic DNA by polymerase
590 chain reaction (PCR) in a Biometra T3000 Thermocycler (30 sec at 98°C ; 30 cycles of
591 10 sec at 98°C, 20 sec at 58°C, 20 sec at 72°C; and elongation for 7 min at 72°C) using
592 the Phusion High-Fidelity DNA Polymerase (NEB, #M0530S) according to the
593 manufacturer's instructions. Primers including Nextera overhangs were used (F-
594 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTGCCCTCCCATTATGTA R-
595 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAGTTCCAAGCCAGCATTTT),
596 allowing library preparation with the Nextera XT Kit (Illumina, #15052163), and
597 sequencing was performed using the Illumina MiSeq system according to manufacturer's
598 instructions. An ESC clone showing a 77 base-pair deletion in exon 4 inactivating *Tal1*
599 gene expression was then injected into C57BL/6 E3.5 blastocysts. Chimeric embryos
600 were subsequently transferred into recipient females at 0.5 days of pseudopregnancy
601 following mating with vasectomised males, as described previously⁴⁴.

602 Chimeric embryos were harvested at E8.5 (7 embryos), dissected, and single-cell
603 suspensions were generated from pooled embryos as described above. Given the low
604 detection rate of the tdTomato transcript (Extended Data Fig. 7b), single-cell suspensions
605 were sorted into tdTomato⁺ and tdTomato⁻ samples using a BD Influx sorter with DAPI at
606 1µg/ml (Sigma) as a viability stain for subsequent 10X scRNA-seq library preparation

607 (version 2 chemistry) and sequencing on an Illumina HiSeq 4000 platform, which resulted
608 in 27,817 tdTomato⁻ and 28,305 tdTomato⁺ cells that passed quality control (see below).
609 Supplementary Information Table 5 contains metadata for each sequenced cell. Flow
610 cytometry of chimeric embryos was performed in parallel using a BD Fortessa cytometer.
611 Cells were stained with the conjugated antibodies CD45-APC-Cy7 (1:200; BD
612 Pharmingen, cat# 557659, clone 30-F11, lot# 6126662), CD41-BV421 (1:200; Biolegend,
613 cat# 133911, clone MWRReg30, lot# B216311), Ter119-PerCP-Cy5 (1:200; Biolegend,
614 cat# 116227, clone TER-119, lot# B169767) and CD71-FITC (1:400; BD Pharmingen,
615 cat# 553266, clone C2, lot# 2307673) with Fc block CD16/32 (1:100; eBioscience, cat#
616 14-0161-85, clone 93, lot# 4316103), and DAPI at 1µg/ml (Sigma) as a viability stain. For
617 the wildtype-into-wildtype experiment, a parental tdTomato⁺ *Tal1*^{+/+} line was injected into
618 C57BL/6 E3.5 blastocysts and processed as for the *Tal1*^{-/-} samples. Three pooled
619 embryos were used for scRNA-seq, and 1,077 tdTomato⁻ and 2,454 tdTomato⁺ cells
620 passed quality control (Supplementary Information Table 6). Chimera sample sizes were
621 dependent on the number of viable embryos that did not show excessive global biases
622 towards host or injected cells (i.e. very low or high fluorescence). Two E9.5 embryos were
623 individually analysed by flow cytometry as described above.

624

625 **10X data pre-processing.** Raw files were processed with *CellRanger* 2.1.1 using default
626 mapping arguments, with reads mapped to the mm10 genome and counted with
627 GRCm38.92 annotation, including tdTomato sequence for chimera cells. HTML reports
628 that provide code, greater detail, and diagnostic plots for the following steps are available
629 at <https://github.com/MarioniLab/EmbryoTimecourse2018>. Singularity containers are

630 also available, providing direct access to the same software versions that were used in
631 this analysis. **Swapped molecule removal.** Molecule counts that derived from barcode
632 swapping were removed from all 10X samples by applying the *DropletUtils* function
633 *swappedDrops* (default parameters) to groups of samples (where a sample is a single
634 lane of a 10X Chromium chip) that were multiplexed for sequencing. **Cell calling.** Cell
635 barcodes that were associated with real cell transcriptomes were identified using
636 *emptyDrops*⁴⁵, which assesses whether the RNA content associated with a cell barcode
637 is statistically significantly distinct from the ambient background RNA present within each
638 sample. A minimum UMI threshold was set at 5,000, and cells with an adjusted p-value <
639 0.01 (BH-corrected) were considered for further analysis. The ambient RNA profile was
640 determined from barcodes associated with fewer than 100 UMIs for the atlas, and 60
641 UMIs for the chimeras. We reproduced our analysis pipeline with a lower UMI threshold
642 of 1,000, and found that no new cell types were present, justifying our rigorous threshold
643 (Extended Data Fig. 2c). **Quality control.** Cell libraries with low complexity (<1000
644 expressed genes) were excluded. Cells with mitochondrial gene expression fractions
645 greater than 2.37%, 2.18%, and 3.35% for each of the wildtype atlas, *Tal1*^{-/-} chimeras,
646 and WT chimeras, respectively, were excluded. The thresholds were determined by
647 considering a median-centred MAD-variance normal distribution; cells with mitochondrial
648 read fraction “outside” of the upper end of this distribution were excluded (adjusted p-
649 value < 0.05; BH-corrected). **Normalisation.** Transcriptome size factors were calculated
650 for each dataset separately (atlas, *Tal1*^{-/-} chimeras, WT chimeras), using
651 *computeSumFactors* from the R *scrn* package⁴⁶. Cells were pre-clustered with *scrn*'s
652 *quickCluster* function (using the *igraph* method), with minimum and maximum cluster

653 sizes of 100 and 3,000 cells, respectively. Raw counts for each cell were divided by their
654 size factors, and the resulting normalised counts were used for further processing.

655 **Selection of highly variable genes (HVGs).** HVGs were calculated using *trendVar* and
656 *decomposeVar* from the *scrn* R package, with loess span 0.05. Genes that had
657 significantly higher variance than the fitted trend (BH-corrected $p < 0.05$) were retained.
658 Genes with mean \log_2 normalised count $< 10^{-3}$, genes on the Y chromosome, *Xist*, and td-
659 Tomato (where applicable) were excluded. **Batch correction.** Batch-effects were
660 removed using the *fastMNN* function in *scrn* on 50 PCs computed from the HVGs only.
661 Correction was performed first between the samples of each time-point, merging
662 sequentially from the samples containing the most cells to the samples containing the
663 least. Time-points were then merged from oldest to youngest; the mixed time-point was
664 merged between E7.25 and E7.0 (Extended Data Fig. 2d). This was applied on the whole
665 atlas dataset, and separately on the subsets of cells considered in Figures 2 and 3 for
666 their respective analyses. Euclidean distances calculated from this batch-corrected PCA
667 were used for all further analysis steps e.g., nearest-neighbour graphs. **Doublet removal.**
668 First, a doublet score was computed for each cell by applying the *doubletCells* function
669 (*scrn* R package) to each 10X sample separately. This function returns the density of
670 simulated doublets around each cell, normalised by the density of observed cell libraries.
671 High scores indicate high doublet probability. We next identified clusters of cells in each
672 sample by computing the first 50 principal components across all genes, building a shared
673 nearest-neighbour graph (10 nearest neighbours; *buildSNNGraph* function; *scrn* R
674 package), and applying the Louvain clustering algorithm (*cluster_louvain* function; *igraph*
675 R package; default parameters) to it. Only HVGs (calculated separately for each sample)

676 were used for the clustering. This procedure was repeated in each identified cluster to
677 break the data into smaller clusters, ensuring that small regions of high doublet density
678 were not clustered with large numbers of singlets. For each cluster, the median doublet
679 score was considered as a summary of the scores of its cells, as clusters with a high
680 median score are likely to contain mostly doublets. Doublet calls were made in each
681 sample by considering a null distribution for the scores of a median-centred MAD-
682 variance normal distribution, separately for each sample. The MAD estimate was
683 calculated only on values above the median to avoid the effects of zero-truncation, as
684 doublet scores cannot be less than zero. All cells in clusters with median score at the
685 extreme upper end of this distribution (BH-corrected $p < 0.1$) were labelled as doublets.
686 A final clustering step was performed across all samples together to identify cells that
687 shared transcriptional profiles with called doublets, but escaped identification in their own
688 samples. Clusters were defined using the same procedure as applied to each sample
689 with the exceptions that sub-clustering was not performed, and batch-corrected principal
690 components were used (see *Batch correction*, above). To identify clusters that contained
691 more doublets than expected, we considered for each cluster the fraction of cell libraries
692 that were called as doublets in their own samples. We modelled a null distribution for this
693 fraction using a median-centred, MAD-estimated variance normal distribution as
694 described for the median doublet score in each sample, above, and called doublets from
695 the distribution as in each sample, above. **Stripped nucleus removal.** Five of the clusters
696 found in the across-sample clustering step above (*Doublet removal*) contained cells with
697 considerably lower mitochondrial gene expression and smaller total UMI counts
698 compared to other clusters. We considered these clusters to consist of nuclei that had

699 been stripped of their cytoplasm in the 10X droplets, and excluded them from downstream
700 analyses due to their supposed technically-derived signal. **Density estimation.** The
701 density of cells in gene expression space was calculated using a tricube kernel on the top
702 50 batch-corrected principal components. The median distance of all cells to their 50th
703 nearest neighbour was used to define the maximum distance for the kernel.

704

705 **Smart-seq2 data pre-processing. Mapping.** Reads were mapped to the mm10 genome
706 using *GSNAP*⁴⁷ (version 2015-09-29) with default arguments except *batch=5*. HTSeq⁴⁸
707 was subsequently used to count the number of reads mapped to each gene using
708 GRCm38.92 for annotation. **Quality control.** Three criteria were used to identify and
709 discard poor-quality cells: (1) Number of mapped reads to nuclear genes < 50,000; (2)
710 number of genes detected < 4,000; (3) proportion of reads mapping to mitochondrial
711 genes > 10%. Cell libraries for which any of these criteria were met were discarded. Of
712 the 288 cell libraries prepared, 250 passed our quality control. **Normalisation.** Cells were
713 size-factor normalised as above (*scrn*).

714

715 **Visualisation. UMAPs** were calculated using *Scanpy* 1.2.2⁴⁹ (*scanpy.api.tl.umap*),
716 considering the 20 nearest neighbours in the batch-corrected PCA, with default
717 parameters except for *min.dist* = 0.7. **Force-directed graphs** considered the 10 nearest
718 neighbours of each cell in a 15-dimension diffusion space calculated on the first 50 PCs
719 of the HVG-subset data (using *Scanpy* v1.2.2 function *tl.diffmap* and *Scanpy* v0.4.4⁴⁹
720 function *utils.comp_distance*). Edges were unweighted, and the layouts were generated
721 in *Gephi* v0.9.2⁵⁰ using the *ForceAtlas2* algorithm⁵¹. **Endoderm diffusion maps** were

722 calculated using the R package *destiny*, with function *DiffusionMap*, using default
723 settings, from batch-corrected PC coordinates. **Graph abstraction**⁵² was computed
724 using the *tl.aga* function from *Scanpy* v1.2.2 and edges were drawn using the adjacency
725 confidence matrix. For Extended Data Fig. 3e-f, *graph abstraction* was computed on the
726 clusters annotated in Fig. 1c and with the threshold for connection of clusters set to 0.23.
727 For Fig. 3b, the clusters in Fig. 3a were used, and the threshold was set to 0.2. For
728 Extended Data Fig. 5a (top), a threshold of 0.95 was used and for Extended Data Fig. 5a
729 (bottom), a threshold of 0.45 was applied. The generation of the clusters for these two
730 latter plots is described in the *Endoderm analysis* section.

731

732 **Clustering and cell annotation.** A shared nearest neighbour graph (considering the 10
733 neighbours of each cell) was constructed using the 50 batch-corrected PCs of the HVG-
734 subset expression data using Euclidean distance (*buildSNNGraph*, R package *scrn*).
735 Clusters were called from this graph using the Louvain algorithm (*cluster_louvain* with
736 default parameters, R package *igraph*). To identify finer substructure from these top-level
737 clusters, each cluster underwent a second round of clustering using the same method as
738 above with the same batch-corrected PC coordinates (i.e., by subsetting from the batch-
739 corrected coordinates). *Graph abstraction* was then used to assess the connectivity
740 across all sub-clusters in the dataset. Within each top-level cluster, we considered
741 distances between the sub-clusters based on their *graph abstraction* connectivity.
742 Specifically, for a connectivity confidence score between clusters of x , we considered a
743 distance of $1-x$. Ward-linkage hierarchical clustering was performed to evaluate sub-
744 cluster relatedness (*scipy.cluster.hierarchy* module, Python 3.4). Sub-clusters with

745 distances less than 1.6 were merged. The merged sub-clusters were then annotated by
746 examination of marker gene expression. Sub-clusters without a unique identity according
747 to marker gene expression were manually merged with their closest sub-cluster to form
748 final cell type annotations. Unannotated sub-clusters (i.e., before merging) are available
749 in Supplementary Information Table 4. **Stability of the atlas under downsampling.** To
750 test stability of the atlas with regard to the size of the dataset, we sampled with
751 replacement cell type labels from the atlas dataset. We performed 50 samplings for each
752 of the sizes of sample (1,000-116,312). For each of these sizes, we calculated for each
753 cell type the ratio of standard deviation of cell type label frequency by mean cell type
754 frequency. The ratios are shown in Extended Data Figure 3d. Note that when the atlas is
755 downsampled to less than half its full size (50,000 cells), the standard deviation of cell-
756 type frequency remains less than 10% of the mean for all cell types.

757

758 **Endoderm analysis. Cell selection.** Cell types annotated as anterior primitive streak,
759 definitive endoderm, visceral endoderm and gut were selected for further analysis. A
760 batch-correction was computed for this cell set, using the same method as described
761 above. **Annotation of the gut.** Cells with the “Gut” cell type label from the collection
762 time-points E8.5 and E8.25 were selected. We constructed a shared nearest-neighbour
763 graph on their batch-corrected PC coordinates (i.e., by subsetting from the coordinates
764 from the endoderm-specific correction; *buildSNNGraph* function; *scrna* R package; 10
765 nearest neighbours), and clustered cells using the *Louvain* algorithm (*cluster_louvain*
766 function; *igraph* R package; default parameters). **Gut tube pseudospacial ordering.**
767 E8.5 cells from the gut clusters (Fig. 2c) were selected and a diffusion map was

768 constructed from their batch-corrected PC coordinates. DPT was calculated for each cell
769 starting from the pharyngeal endoderm cell with minimum value on DC2. **Differential**
770 **expression analyses** were performed using the *findMarkers* function in *scrn*, using the
771 10X sample as a blocking factor. Significantly differentially expressed genes were
772 considered as those with BH-corrected $p < 0.1$. Hindgut differentially expressed genes
773 were tested against an absolute fold-change of 0.5. **Transport maps**¹⁶. This approach
774 considers each cell as a unit of mass, to be “transported” to other cells at consecutive
775 time-points. By seeking to move these masses efficiently between time-points (i.e.,
776 minimising the transcriptional differences between cells across which mass is moved) a
777 mapping of expected descendant and ancestor cells can be identified. Importantly, this
778 method allows the integration of both transcriptional and collection time-point information.
779 Transport maps were constructed using the *wot* python package (v0.2.1) using default
780 settings except for skipping the dimension-reduction step, and instead using the batch-
781 corrected PCs as input. 100 randomly selected cells from each collection time-point of
782 ExE endoderm were added to the cells projected in Fig. 2c (Extended Data Fig. 4h). Cells
783 from the mixed time-points were excluded from the analysis. **Selecting cells for**
784 **trajectories with the transport maps**. For pushing mass forward through the graph (i.e.,
785 considering from which progenitor cells the gut clusters derived), we considered two
786 starting populations. The DE population consisted of E7.0 cells labelled as anterior
787 primitive streak or definitive endoderm; the VE population consisted of E7.0 cells labelled
788 as visceral endoderm. This stage was selected because the two populations still retained
789 very distinct transcriptional profiles, and many cells were present for each population. For
790 pulling mass backward through the graph (i.e., selecting cells for the DE-Hindgut 2 and

791 VE-Hindgut 1 trajectories), we considered E8.5 cells from each of the gut clusters as
792 terminal populations. For the cells in the VE-Hindgut 1 trajectory, we included all cells
793 whose largest mass contribution was to the Hindgut 1 cluster. For cells in the DE-Hindgut
794 2 trajectory, we selected cells whose Hindgut 2 mass contribution was greater than 90%
795 their largest mass contribution to any cluster. This allows selection of cells committed to
796 Hindgut 2 (i.e., with greatest mass towards Hindgut 2), and common progenitor cells,
797 which show relatively balanced mass contributions to several terminal clusters. Cells with
798 balanced mass contributions across clusters were not observed for the VE-Hindgut 1
799 trajectory, consistent with the hindward bias of the intercalated visceral endoderm cells.

800 **Selecting cells for trajectories with graph abstraction.** A shared 10-nearest neighbour
801 graph was constructed on the endoderm cell subset using the first 50 PCs of the HVG-
802 subset expression data using Euclidean distance (*buildSNNGraph*, R package *scrn*).
803 Clusters were called from this graph using the Louvain algorithm (*cluster_louvain* with
804 default parameters, R package *igraph*). Clusters that presented more substructure in the
805 force-directed layout were further subclustered using the same pipeline. Cells were
806 selected based on a manually curated parsimonious trajectory connecting Hindgut 1 or 2
807 to the appropriate progenitor populations (Extended Data Fig. 5a, b). **Hindgut**
808 **trajectories and gut pseudospace ordering.** Genes were clustered as in⁵ with some
809 modifications. First, cells were ordered using DPT, calculated from a diffusion map built
810 from the endoderm-specific batch-corrected PC coordinates of the relevant subset of
811 cells. DPT was calculated from a cell with the most extreme value on DC1; direction along
812 DC1 was selected to start from the youngest populations of cells. HVGs were calculated
813 for each cell subset, with only these genes retained for subsequent clustering. For each

814 retained gene, we fitted two ordinary least squares linear models (constant and degree-
815 2 polynomial functions) that regress the \log_2 normalised expression levels for each cell
816 against the values of DPT calculated above. Genes for which the degree-2 polynomial fit
817 the data better were retained (F-test, BH-corrected $p < 0.05$, R function *anova.lm*). For
818 each of these genes, we fitted a local regression to the expression level for each cell at
819 their value of DPT (R function *loess*, span = 0.75). We then identified the predicted value
820 of the loess fit for one thousand uniformly spaced points across the DPT to provide
821 smoothed gene expression estimates and avoid biasing clustering to regions of DPT with
822 high cell density. loess fits were scaled to a range of (0,1) to prevent clustering by
823 expression level. The Pearson correlation distance between each gene was calculated
824 as $([1-x]/2)^{0.5}$, where x is the Pearson correlation, and hierarchical clustering (UPGMA)
825 was performed. The tree was cut with *dynamicTreeCut* (R; minimum cluster size of 50
826 genes, otherwise default parameters).

827

828 **Blood development analysis. Cell clustering.** A 10-nearest neighbour graph was
829 constructed on the haemato-endothelial cell subset using the first 50 PCs of the HVG-
830 subset expression data and Euclidean distance (*buildKNNGraph*, R package *scrn*).
831 Clusters were called from this graph using the Louvain algorithm (*cluster_louvain* with
832 default parameters, R package *igraph*). Two clusters that presented higher substructure
833 in the force-directed layout (one in EC, containing EC3-8; and one in the Haem/BP region,
834 containing Haem3-4, BP3-4, My and Mk) were further subclustered using the same
835 pipeline but different k values: $k=30$ for EC cluster and $k=15$ for Haem/BP cluster.
836 **Differential expression analyses.** Pairwise comparisons were performed using

837 edgeR⁵³. Dispersions were estimated using *estimateCommonDisp* and
838 *estimateTagwiseDisp*, tests with *exactTest* function and p-values BH-corrected. All
839 functions were used with default parameters. **Mapping of Smart-seq2 data to the**
840 **reference atlas.** The Spearman correlation distance, $([1-x]/2)^{0.5}$, was computed between
841 each cell in the Smart-seq2 dataset and each cell in the endothelium cluster from the 10X
842 atlas using the HVGs computed for the EC clusters of Fig. 3. The labels of the atlas
843 endothelial cells were defined as the most frequent dissection location within the 5
844 nearest neighbours. If cells had an equal number of neighbours from two locations, they
845 remained unassigned. **Mapping of published embryonic blood dataset⁵⁴.** To support
846 the annotation of the myeloid cluster, atlas cells from Fig. 3a were mapped to a published
847 dataset containing haematopoietic cells collected between E9.5 and E11.5⁵⁴ (Extended
848 Data Fig. 6g). The mapping was performed as with the Smart-seq2 dataset, using the
849 HVGs computed for the published dataset⁵⁴. The published dataset was processed as
850 follows: The counts matrix with transcript counts per million (TPM) was downloaded from
851 GEO accession GSE87038. Counts were log transformed as $\log_2(n/10 + 1)$ where n is
852 the TPM value, and HVGs were calculated as in⁵⁵. Since cluster identities from ⁵⁴ were
853 not provided, the data was re-clustered using *Louvain* clustering on a k-nearest neighbour
854 graph with $k=10$ considering only the HVGs. Clusters were subsequently merged to
855 approximate the clusters and expression patterns of marker genes shown in Figure 8 of⁵⁴.

856

857 **Tal1^{-/-} chimera analysis. Mapping to the atlas.** To avoid mapping biases that derive
858 from unequal numbers of atlas cells at each collection time-point, each stage of the atlas
859 was sub-sampled at random such that 10,000 cell libraries (i.e., including doublets and

860 stripped nuclei) were present at each time-point. Cells from the mixed time-point were
861 excluded. Stages E6.5 and E6.75 contained fewer cells (3,697 and 2,169 respectively)
862 and were not downsampled; however, we do not expect cells from E8.5 chimera to map
863 to these time-points, so their cell number bias is likely to be unimportant. We first
864 constructed a 50-dimensional PC space from the combined normalised log-counts of
865 subsampled atlas cells (including doublets and stripped nuclei) and the cells from the
866 samples that are to be mapped to the atlas. Batch-correction was then performed on the
867 atlas cells in the PC space, as described above (*Batch correction*), to construct a single
868 reference manifold for mapping. Samples to be mapped were then independently merged
869 with the newly-corrected atlas data (*scrn* function *fastMNN*), and the 10 nearest cells
870 (Euclidean distance) in the atlas to each chimera cell were recorded. Mapped time-point
871 and cell type of chimera cells were defined as the mode of those of its nearest-neighbours.
872 Ties were broken by choosing the stage or cell type of the cell that had the lowest distance
873 to the chimera cell. Cells that mapped to doublet- or stripped nucleus-labelled cells were
874 excluded from downstream analyses. The robustness of this mapping was assessed by
875 mapping one entire biological replicate of E8.0 cells onto the atlas, having removed these
876 cells from the reference. 89.4% of these cells correctly mapped to their annotated cell
877 type (Extended Data Fig. 7c), and 29.2% to the correct time-point (Extended Data Fig.
878 7d; 83.1% of cells mapped within one time-point in either direction). **Visualisation.** UMAP
879 visualisation of the data was performed as described above, using 50 batch-corrected
880 PCs. Batches of cells of the same genotype were merged first, followed by merging
881 across genotypes. To show the cell mapping with respect to atlas landscapes (e.g. Fig.
882 4b), we coloured cells in the visualisation by the closest atlas cell for each chimera cell

883 after mapping. **Remapping of cells to the haemato-endothelial landscape.** To ensure
884 that we utilised the full resolution of our atlas, a subset of chimera cells were mapped to
885 the complete (i.e., not subsampled) atlas dataset for the relevant cell types. The mapping
886 procedure was repeated as described above, only for atlas and chimera cells from the
887 erythroid, haemato-endothelial, blood progenitor, endothelial and mixed mesoderm cell
888 types. No downsampling was performed. **Relative contributions to atlas clusters from**
889 **injected and host cells in the *Tal1*^{-/-} chimera.** To compare the frequency of cells that
890 contribute from each of the host and injected populations of cells to atlas clusters, we first
891 corrected for compositional differences between the populations: ExE endoderm, Visceral
892 endoderm, ExE ectoderm, Parietal endoderm, Blood Progenitors 1-3 and Erythroid 1-3
893 were excluded from this analysis, as the injected (*Tal1*^{-/-}) cells cannot contribute to these
894 lineages. The frequency of cells from each subcluster was then calculated and log-fold
895 change calculated (Fig. 4e). Due to the absence of cells in the *Tal1*^{-/-} samples, BP and
896 Ery sub-clusters were not considered. **Differential expression analyses** were performed
897 as in the “Blood development analysis” section.

898

899 **Additional References**

- 900 39 Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and
901 ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
- 902 40 Nichols, J. & Jones, K. Derivation of Mouse Embryonic Stem (ES) Cell Lines Using
903 Small-Molecule Inhibitors of Erk and Gsk3 Signaling (2i). *Cold Spring Harbor*
904 *Protocols* **2017**, pdb.prot094086 (2017).

905 41 Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**,
906 519-523 (2008).

907 42 Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression
908 of the pluripotency network and increases embryonic stem cell resistance to
909 differentiation. *Nat. Cell Biol.* **13**, 838-845, doi:10.1038/ncb2267 (2011).

910 43 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat.*
911 *Protoc.* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

912 44 Bin, G. C. L. *et al.* Oct4 is required for lineage priming in the developing inner cell
913 mass of the mouse blastocyst. *Development* **141**, 1001-1010 (2014).

914 45 Lun, A. *et al.* Distinguishing cells from empty droplets in droplet-based single-cell
915 RNA sequencing data. *bioRxiv*, 234872 (2018).

916 46 Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level
917 analysis of single-cell RNA-seq data. *F1000Research* **5**, 2122 (2016).

918 47 Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and
919 splicing in short reads. *Bioinformatics* **26**, 873-881 (2010).

920 48 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
921 throughput sequencing data. *Bioinformatics* **31**, 166-169,
922 doi:10.1093/bioinformatics/btu638 (2015).

923 49 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene
924 expression data analysis. *Genome Biol.* **19**, 15 (2018).

925 50 Bastian, M., Heymann, S. & Jacomy, M. in *Third International AAAI Conference on*
926 *Weblogs and Social Media.*

- 927 51 Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. ForceAtlas2, a Continuous
928 Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi
929 Software. *PLoS One* **9**, e98679 (2014).
- 930 52 Wolf, F. A. *et al.* Graph abstraction reconciles clustering with trajectory inference
931 through a topology preserving map of single cells. *bioRxiv*, 208819 (2017).
- 932 53 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package
933 for differential expression analysis of digital gene expression data. *Bioinformatics*
934 **26**, 139-140 (2010).
- 935 54 Dong, J. *et al.* Single-cell RNA-seq analysis unveils a prevalent
936 epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol.*
937 **19**, 31 (2018).
- 938 55 Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq
939 experiments. *Nature Methods* **10**, 1093-1095 (2013).
- 940 56 Kinder, S. J. *et al.* The orderly allocation of mesodermal cells to the extraembryonic
941 structures and the anteroposterior axis during gastrulation of the mouse embryo.
942 *Development* **126**, 4691-4701 (1999).
- 943 57 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells.
944 *Nat Commun* **8**, 14049 (2017).

945

946 **Data availability statement**

947 Raw sequencing data is available on ArrayExpress: Atlas – E-MTAB-6967; Smart-seq2
948 endothelial cells – E-MTAB-6970; *Tal1*^{-/-} chimeras – E-MTAB-7325; WT chimeras – E-

949 MTAB-7324. Processed data may be downloaded following the instructions at
950 <https://github.com/MarioniLab/EmbryoTimecourse2018>. GEO accession GSE87038 was
951 used to support the annotation of myeloid cells (see Methods). All code is available upon
952 request, and at <https://github.com/MarioniLab/EmbryoTimecourse2018>. Our atlas can be
953 explored at <https://marionilab.cruk.cam.ac.uk/MouseGastrulation2018/>

954

955 **Extended Data Figure Legends**

956 **Extended Data Figure 1: Embryo images.** Representative images of embryos collected
957 at the time-points indicated in the Figure. Scale bars: 0.25mm.

958

959 **Extended Data Figure 2: Data quality control.** **a**, Left: Estimated number of cells
960 present in a single mouse embryo at each time-point. Points are values measured in²⁴;
961 line is an Ordinary Least Squares regression fit. Right: Number of cells captured in this
962 study compared to the number of cells estimated in the embryo in (Left). **b**, Violin plots
963 illustrating the number of detected genes (top) and Unique Molecular Identifiers (UMIs,
964 bottom) per cell per sample. Sample 11 failed quality control and is therefore not shown.
965 Sample details are provided in Supplementary Information Table 1. **c**, UMAP highlighting
966 additional cells identified considering a reduced UMI threshold of 1,000. Additional cells
967 are shown in black. Cells from the atlas are shown in the colour corresponding to their
968 cell type (Fig. 1c). Note that all additional cells are present alongside cells from the atlas:
969 no new cell types are found. **d**, UMAP as shown in Fig. 1c with cells coloured by biological
970 replicate, showing consistency between samples collected at the same time-point. **e**,

971 Mean gene expression of diagnostic markers (y-axis) for each cell type (x-axis) are
972 shown. Genes are row-normalised.

973

974 **Extended Data Figure 3: Lineage progression.** **a**, UMAP as shown in Fig. 1c, coloured
975 by the density of each cell in gene expression space; brighter coloured regions (towards
976 yellow) are more densely sampled, while darker regions (towards blue) are more sparsely
977 sampled relative to other regions in the atlas. **b**, Boxplots summarising the density per
978 cell type. Values are log-transformed. **c**, UMAPs as shown in Fig. 1c, highlighting cells
979 from each sampled time-point illustrating the transcriptional progression along
980 developmental time. **d**, Results of atlas stability testing (see Methods). y-axis: ratio of
981 standard deviation of cell type frequency by the mean cell-type frequency at different
982 degrees of downsampling. Note that when the atlas is downsampled to less than half its
983 full size (50,000 cells), the standard deviation remains less than 10% of the mean for all
984 cell types. **e, f**, *Abstracted graphs*, which quantify the degree of similarity between the
985 identified clusters to represent the underlying biological structure of the dataset. Nodes
986 correspond to the annotated cell types, and edges reflect the confidence of adjacency
987 between clusters (thicker edges indicate higher confidence). Node sizes increase as a
988 function of the number of cells within each cluster. Nodes in **(e)** are coloured and
989 numbered according to the legend shown in Fig. 1c. Nodes in **(f)** show the frequency of
990 cells from each time-point, excluding two samples of mixed time-point embryos.

991 **Extended Data Figure 4: Endoderm convergence.** **a**, Schematic representing the
992 process of definitive endoderm intercalation following gastrulation, and subsequent gut
993 maturation. Adapted from⁹. **b-g**, Gene expression levels of *Ttr* (b), *Mixl1* (c), *Nkx2-5* (d),

994 *Pyy* (e), *Nepn* (f), *Cdx2* (g) overlaid on the Fig. 2a Force-directed graph. **h**, Diffusion map
995 of cells selected for transport map construction; cells selected as termini for pulling mass
996 backward through the transport maps are coloured. **i**, Results of pushing mass forward
997 through the transport maps are shown on the force-directed layout. **j**, Violin plots showing
998 expression levels of *Trap1a* and *Rhox5* in all cell-types of the full atlas. **k**, Dorsal view of
999 a whole-mount fluorescence image of a *Ttr::Cre*; R26R::YFP embryo at E8.5. Green:
1000 YFP, Red: Phalloidin. Arrowhead: increased Ttr-YFP staining in the posterior region of
1001 the gut. Scale bar: 300 μ m.

1002 **Extended Data Figure 5: Endoderm trajectories.** **a**, Top: *Graph abstraction* of the
1003 endoderm landscape after fine sub-clustering as an alternative method to resolve which
1004 cells should be part of the VE-Hindgut 1 trajectory or the DE-Hindgut 2 trajectory
1005 (supporting transport maps; see Methods). Edges along VE-Hindgut 1 trajectory
1006 highlighted in yellow (nodes 1-4; yellow numbers). Edges along DE-Hindgut 2 trajectory
1007 highlighted in orange (nodes 1-12; orange numbers). Bottom: *Graph abstraction* with the
1008 subset of nodes related to the DE-Hindgut 2 trajectory to resolve the origin of cluster 4
1009 (between 5 and 6 in top panel). Resulting DE-Hindgut2 trajectory includes clusters 1-4
1010 and 6-9. The right hand panel overlays information about the composition of each cluster
1011 by developmental stage. **b**, Force-directed graph coloured by *graph abstraction* (PAGA)
1012 trajectories. Note that this independent approach for trajectory identification reaches very
1013 similar results to those inferred by *transport maps* in Fig. 2h. **c**, Gene-normalised
1014 dynamics of all clusters found along the VE-Hindgut 1 and the DE-Hindgut 2 trajectory (x-
1015 axis: DPT along the trajectory, y-axis: normalized expression levels). The black line is the
1016 mean fitted expression level across all genes in each cluster. Grey shading indicates the

1017 standard deviation along the trend across all genes in the cluster. Pink area highlights
1018 intercalation process; blue area highlights gut maturation steps. Dashed lines correspond
1019 to additional stages in the process deduced from the changes in gene expression trends.
1020 Points below the plots are the DPT coordinates of cells from each time-point coloured
1021 according to time-point as in Fig. 1f (from E6.5 in red to E8.5 in blue). **d**, Gene-normalised
1022 dynamics of VE genes along VE-Hindgut 1 trajectory indicating VE maturation prior to the
1023 intercalation stage. Plot design is as in (c); **d-g**: Below the x-axis, points are as in (c). **e**,
1024 Left: Venn diagram of genes up-regulated during the intercalation process in both VE-
1025 Hindgut 1 (in clusters 3, 5, 8, 11) and DE-Hindgut 2 (in clusters 4, 6, 7, 8, 9) trajectories.
1026 Listed genes: signature of epithelial remodelling in the overlapping fraction. Right:
1027 expression trends of illustrative genes along the trajectories (gene-normalised). **f**, Left:
1028 Venn diagram of genes up-regulated after the intercalation process in both trajectories
1029 (VE-Hindgut 1: clusters 1, 2, 5, 10; DE-Hindgut 2: clusters 1, 3, 5, 10); the overlapping
1030 fraction was enriched in transcription factors including a large subset of homeodomain
1031 proteins (listed). Right: gene-normalised dynamics of *Hox* and *Cdx* genes along the
1032 trajectories. **g**, Gene-normalised dynamics of transcription factors up-regulated
1033 specifically in the VE-Hingut 1 trajectory during endoderm intercalation.

1034 **Extended Data Figure 6: Blood development.** **a**, Diagram illustrating embryonic blood
1035 emergence from the two first waves. At E6.5, gastrulation begins. Using transplantation
1036 assays, it has been shown that the proximo-posterior epiblast cells closest to the primitive
1037 streak (PS) at this stage (red) mainly give rise to primitive erythroid cells in the YS, while
1038 the epiblast cells located in the middle of the embryo at E6.5 but closer to the PS at a
1039 later stage are enriched for endothelial progenitors⁵⁶. At E7.5, blood islands are apparent

1040 (zoomed box of primitive blood wave), where primitive erythroid cells are surrounded by
1041 endothelium. At around E8.25, some endothelial cells (haemogenic endothelium)
1042 undergo an endothelial-to-haematopoietic transition and become Erythroid-Myeloid
1043 Progenitors (EMPs), which migrate to the foetal liver (FL) and give rise to definitive
1044 erythrocytes. Adapted from³⁸. **b**, Force directed layout of Fig. 3a coloured by original
1045 clusters from Fig. 1c. **c**, Gene expression levels of *Cdh5* and *Pecam1* overlaid on the
1046 *graph abstraction* visualisation from Fig. 3b. **d**, Experimental design to isolate FLK1⁺ cells
1047 from yolk sac, allantois and embryo proper for Smart-seq2 scRNA-seq. **e**, Representative
1048 image of an embryo collected for the transcriptional analysis of endothelial cells from the
1049 yolk sac (YS), allantois (AL) and Embryo proper (EP). **f**, Sorting strategy of FLK1⁺ cells
1050 from YS, EP and AL on live cells (DAPI⁻). x-axis: FLK1 intensity. y-axis: DAPI intensity.
1051 **g**, Evidence to support myeloid annotation of My cluster in Fig. 3. Haemato-endothelial
1052 cells from Fig. 3a were mapped to a published dataset that profiled haematopoietic cells
1053 collected at E9.5, E10.5 and E11.5 from different organs⁵⁴. Barplots show the fraction of
1054 atlas cells in the My cluster mapped to the clusters defined by Dong et al.⁵⁴ Figure 8. **h**,
1055 Representative images of the dissected regions collected to study the location of
1056 CSF1R⁺CD16/32⁺ cells. Scale bar: 0.25mm. **i**, Flow cytometry plots indicating the
1057 frequency of CSF1R⁺CS16/32⁺ cells in each embryonic region. Two biological replicates
1058 were performed for this experiment: with pools of 12 and 13 embryos respectively. Plots
1059 illustrate one biological replicate.

1060

1061 **Extended Data Figure 7: Analysis of *Tal1*^{-/-} chimeras.** **a**, Representative chimera
1062 embryo harvested at E8.5 (left: brightfield, right: tdTomato fluorescence; scale bar:

1063 0.25mm), and flow cytometry plot with tdTomato fluorescence distribution and sorting
1064 gates. **b**, Histograms showing the UMI counts for the tdTomato construct in both
1065 tdTomato⁻ and tdTomato⁺ fractions in the *Tal1*^{-/-} into WT experiment (see Methods). **c-d**,
1066 Control mapping results of an E8.0 biological replicate that was removed and mapped
1067 back to the atlas. **c**, Heatmap showing the fraction of cells of each labelled cell type that
1068 mapped to each cell type in the reference atlas. Numbers above columns indicate the
1069 number of cells in each category. 89.4% of these cells correctly mapped to their annotated
1070 cell type. **d**, Histogram showing the fraction of cells from the E8.0 biological replicate that
1071 mapped to each time-point in the reference. 29.2% of cells mapped to the correct time-
1072 point, 83.1% of cells mapped within one time-point in either direction. **e**, Scatter plot
1073 comparing the percentage of tdTomato⁺ cells against tdTomato⁻ for each cell type in both
1074 *Tal1*^{-/-} into WT (left) and WT into WT (right) experiments. Black arrow: extra-embryonic
1075 tissues. White arrow: haematopoietic tissues. **f**, Force-directed graph of blood-related
1076 lineages from the atlas (Fig. 3), coloured by *Tal1* expression levels. Darker colouring
1077 shows higher expression. **g**, Flow cytometry analysis of E8.5 *Tal1*^{-/-} into WT chimeras
1078 showing the complete depletion of the haematopoietic markers CD41, CD45 (upper), as
1079 well as of the CD71⁺ Ter119⁺ erythroid fraction (lower) in *Tal1*^{-/-} tdTomato⁺ cells (right
1080 panels). **h**, UMAPs of WT into WT experiment showing balanced contribution to all
1081 embryonic lineages. **i**, Flow cytometry analysis of WT into WT chimeras showing
1082 balanced contribution to the haematopoietic lineage from both tdTomato⁺ and tdTomato⁻
1083 cells at E9.5 (representative of 2 individual embryos).

1084 **Extended Data Figure 8: Transcriptional effects of disruption caused by *Tal1*.** **a**,
1085 Heatmap illustrating the row-normalised expression of genes upregulated in EC3-

1086 mapped *Tal1*^{-/-} cells when compared to their closest neighbours in the atlas (labelled
1087 “EC3”) and EC3-mapped WT chimera cells (labelled “WT”). **b**, UMAPs as in Fig. 1c,
1088 showing the expression of *Tdo2*, *Plagl1* and *Pcolce*.

1089