1   **Single-cell transcriptome maps of myeloid blood cell lineages in *Drosophila***

2

3   Bumsik Cho[1][#], Sang-Ho Yoon[1][#], Daewon Lee[1], Ferdinand Koranteng[1], Sudhir

4   Gopal Tattikota[2], Nuri Cha[1], Mingyu Shin[1], Hobin Do[1], Yanhui Hu[2], Sue Young Oh[3],

5   Seok Jun Moon[3], Norbert Perrimon[2,4], Jin-Wu Nam[1,5,6]* and Jiwon Shim[1,5,6]*†

6

7   [1]Department of Life Sciences, College of Natural Science, Hanyang University,

8   Seoul 04736, Republic of Korea

9   [2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

10   [3]Department of Oral Biology, Yonsei University, College of Dentistry, Seoul

11   03722, Republic of Korea

12   [4]Howard Hughes Medical Institute, Boston, MA 02115, USA

13   [5]Research Institute for Natural Sciences, Hanyang University, Seoul 04736,

14   Republic of Korea

15   [6]Research Institute for Convergence of Basic Sciences, Hanyang University, Seoul

16   04736, Republic of Korea

17

18   [#]These authors contributed equally to the work

19   [†]Lead contact

20   * Corresponding authors: jshim@hanyang.ac.kr; jwnam@hanyang.ac.kr

21
22

23   **Key words**: *Drosophila*, hemocyte, myeloid blood, lymph gland, single-cell RNA

24   sequencing, wasp infestation, hematopoietic stem cell, plasmatocyte, crystal cell,

25   lamellocyte, hematopoiesis, innate immunity, Seurat, Drop-seq

26 **SUMMARY**

27   *Drosophila* lymph gland, the larval hematopoietic organ comprised of

28   prohemocytes and hemocytes, has been a valuable model for understanding

29   mechanisms underlying hematopoiesis and immunity. Three types of mature

30   hemocytes have been characterized in the lymph gland: plasmatocytes,

31   lamellocytes, and crystal cells, which are analogous to vertebrate myeloid cells.

32   Here, we used single-cell RNA sequencing to comprehensively analyze

33   heterogeneity of developing hemocytes in the lymph gland, and discovered novel

34   hemocyte types, stem-like prohemocytes, and intermediate prohemocytes.

35   Additionally, we identified the emergence of the lamellocyte lineage following

36   active cellular immunity caused by wasp infestation. We unraveled similarities

37   and differences between embryonically derived- and larval lymph gland

38   hemocytes. Finally, the comparison of *Drosophila* lymph gland hemocytes and

39   human immune cells highlights similarities between prohemocytes and

40   hematopoietic stem cell, and between mature hemocytes and myeloid cells

41   across species. Altogether, our study provides detailed insights on the

42   development and evolution of hematopoiesis at single-cell resolution.

43   **INTRODUCTION**

44   Blood cells are highly specialized cells that play crucial roles such as

45   elimination of foreign threats during immune responses and retaining memories

46   of immunological events (Sakaguchi et al., 2010). In vertebrates, the multifaceted

47   immune system is composed of two lineages, phagocytic myeloid and memory-

48   dependent lymphoid lineages, to allow holistic and cooperative defense of

49   animals (Weissman et al., 2001). Blood cells in *Drosophila*, collectively called

50   hemocytes, are reminiscent of myeloid-lineage blood cells in vertebrates

51   (Banerjee et al., 2019; Crozatier and Vincent, 2011; Gold and Bruckner, 2014),

52   and are represented by at least three morphologically distinct hemocyte

53   populations: plasmatocytes (PM), crystal cells (CC), and lamellocytes (LM).

54   Plasmatocytes, which comprise ~95 % of the hemocytes, play a role in

55   phagocytosis, tissue remodeling, and cellular immune responses – much like

56   macrophages, their vertebrate counterparts (Franc et al., 1996; Irving et al.,

57   2005; Kocks et al., 2005; Kurucz et al., 2007). Crystal cells account for ~5 % of

58   the blood population and are characterized by crystalline inclusions made up of

59   prophenoloxidase (ProPO), an enzyme responsible for activating the

60   melanization cascade (Binggeli et al., 2014; Gajewski et al., 2007; Lebestky et al.,

61   2000; Tang et al., 2006). Finally, lamellocytes, which are seldom seen in healthy

62   animals grown at normal conditions, mostly differentiate upon parasitic wasp

63   infestation or environmental challenges (Anderl et al., 2016; Honti et al., 2014;

64   Rizki and Rizki, 1984; Sorrentino et al., 2007; Xavier and Williams, 2011).

65   Blood development in vertebrates involves the primitive and definitive

66   waves of hematopoiesis (Galloway and Zon, 2003). Reminiscent of vertebrate

67   hematopoiesis, two hematopoietic waves have been described during *Drosophila*

3

68   development, embryonic and larval definitive hematopoiesis (Evans et al., 2003;

69   Hartenstein, 2006). Embryonic hematopoiesis initiates in the head mesoderm of

70   stage-7-embryos and gives rise to hemocytes that migrate throughout the

71   embryo (Holz et al., 2003; Tepass et al., 1994). Upon hatching, embryonically

72   derived hemocytes spread throughout the larva with one hemocyte population

73   freely circulating in the hemolymph and the other colonizing local

74   microenvironments including segmentally repeated hematopoietic pockets of

75   the larval body wall (Leitao and Sucena, 2015; Makhijani et al., 2011).

76   Definitive hematopoiesis is initiated from hemangioblast-like cells in the

77   embryonic cardiogenic mesoderm, which give rise to the larval lymph gland

78   (Mandal et al., 2004). Medially located prohemocytes, which sustain the

79   developmental potential to give rise to all three mature hemocyte types,

80   constititute the medullary zone (MZ) and continue to proliferate until the early

81   third instar (Jung et al., 2005). Mature hemocytes emerge at the distal edge of the

82   lymph gland, the cortical zone (CZ), from mid-second instar (Kocks et al., 2005;

83   Kroeger et al., 2012). Located between the undifferentiated MZ and the

84   differentiated CZ, is the intermediate zone (IZ) that contains a group of

85   differentiating cells expressing markers for both the MZ and the CZ (Krzemien et

86   al., 2010). The posterior signaling center (PSC), a small group of cells that secrete

87   various ligands, regulates proper growth of the rest of the lymph gland

88   (Benmimoun et al., 2015; Krzemien et al., 2007; Mandal et al., 2007). Lymph

89   glands from healthy larvae reared under normal lab conditions generally follow

90   fixed developmental states until late third instar. Remarkably, following the

91   onset of pupariation, the lymph gland disintegrates, allowing hemocytes to

92   disperse into circulation (Grigorian et al., 2011; Regan et al., 2013).

93   Female wasps attack second-instar larvae via a sharp needle-like

94   ovipositor that efficiently delivers their eggs (Lemaitre and Hoffmann, 2007;

95   Yang et al., 2015). Wasp eggs trigger cellular immune responses that accompany

96   lamellocyte differentiation in both embryonic and lymph gland hemocytes.

97   Lamellocytes are seen in circulation by 24 hours post-infestation, whereas lymph

98   gland hemocytes remain intact in their original location. Within 48 hours after

99   infestation, a massive differentiation of lamellocytes takes place followed by

100   disruption of the lymph gland (Markus et al., 2009; Sorrentino et al., 2002;

101   Tokusumi et al., 2009). Hemocytes in the lymph gland eventually dissociate into

102   circulation, and mature lamellocytes derived from the lymph gland and

103   hematopoietic pockets encapsulate and neutralize wasp eggs (Irving et al., 2005;

104   Lanot et al., 2001).

105   The *Drosophila* lymph gland has been largely characterized based on

106   genetic markers and cellular morphology. However, the molecular

107   underpinnings of hematopoietic cells such as different states and the gene

108   regulatory network of each cell type have been less investigated. In addition,

109   questions as to how prohemocytes and mature hemocytes differentiate into

110   lamellocytes upon active immunity, and to what extent hemocytes derived from

111   the embryonic and the lymph gland hematopoiesis differ have been unanswered.

112   Furthermore, due to the lack of sufficient molecular and genetic information, the

113   similarity between *Drosophila* hemocytes and vertebrate immune cells remains

114   to be clarified.

115   Here, we build an atlas of myeloid-like *Drosophila* hemocytes by taking

116   advantage of single-cell RNA sequencing (scRNA-seq) technology and establish a

117   detailed map for larval hemocytes in the developing lymph gland. We uncover

118   novel classes of hemocytes and their differentiation trajectories, and describe

119   molecular and cellular changes of myeloid hemocytes upon immune challenges.

120   Furthermore, we identify both distinct and common characteristics of hemocytes

121   originating from embryonic and definitive lineages. We also document the

122   evolutionary similarities between *Drosophila* hemocytes and human immune

123   cells. Altogether, our study will stimulate future studies on the function and

124   evolution of the myeloid blood cell lineage.

125 **RESULTS**

126 **Single-cell transcriptomic profiling of developing myeloid hemocytes**

127   The lymph gland is the larval hematopoietic organ composed of multiple

128 hemocyte cell types and states (Banerjee et al., 2019; Crozatier and Vincent,

129 2011). To understand the cellular diversity of developing myeloid-like

130 hemocytes in *Drosophila* lymph glands at a single-cell level, we dissected and

131 dissociated lymph glands at three developmental time points, 72, 96, and 120

132 hours after egg laying (AEL), and applied single cells to Drop-seq, a droplet-

133 based microfluidics platform (Macosko et al., 2015) (Figure 1A). Fourteen

134 independent sequencing libraries, representing 5 each for 72 and 96 h AEL and 4

135 for 120 h AEL, were prepared for scRNA-seq. We integrated the sequencing

136 libraries after correcting for batch effects within and between time points using

137 Seurat3 (Butler et al., 2018; Stuart et al., 2019). Our quality-control pipeline

138 eliminated cells with outlier unique molecular barcode (UMI) counts, low gene

139 numbers, high mitochondrial gene contents as well as doublets predicted by

140 Scrublet (see Methods for details). As a result, a total of 22,645 cells (72 h AEL:

141 2,321; 96 h AEL: 9,400; 120 h AEL: 10,924 cells) were retained for subsequent

142 analyses (Figure S1A). The number of cells yielded 5.5 X, 6.8 X, and 2.4 X cell

143 coverage of one lymph gland lobe at 72, 96, and 120 h AEL, respectively (Figure

144 1B). The 22,645 high-quality cells exhibited a median of 6,361 transcripts (UMIs)

145 and 1,477 genes per cell (Table S1; Figure S1B-S1C). In addition, the scRNA-seq

146 libraries of individual time points corresponded well with genes detected in bulk

147 RNA-seq ($\geq$ 1 TPM; 8,724, 7,654, and 7,627 genes at 72, 96, and 120 h AEL,

148 respectively), while undetected remainders displayed low expression levels

149 (Figure S1D). Furthermore, we validated that the scRNA-seq libraries from 72,

150   96, or 120 h AEL align well with the pseudotime array of each library (Figure

151   S1E). Altogether, our libraries appear sufficiently complex to reflect the whole

152   transcriptome of developing hemocytes including minor cell types.

153

154   **Major cell types and transcriptional dynamics of *Drosophila* hematopoiesis**

155   After validating the quality of the single cell libraries, we mapped the cells

156   to the major zones of the lymph gland (Jung et al., 2005). To explore the major

157   cell types in the developing lymph gland, we aligned cell clusters from the three

158   developmental time points using the Louvain algorithm (Blondel et al., 2008)

159   and visualized the data using a nonlinear dimensionality reduction by *t*-

160   distributed stochastic neighbor embedding (*t*-SNE) (van der Maaten and Hinton,

161   2008). We aggregated cell clusters according to the expression of previously

162   published marker genes by manual curation and identified seven distinctive

163   groups of isolated hemocytes (Figure 1C; Figure S1F-1G; Table S2). These

164   clusters include prohemocytes (PH: *Tep4*, *Ance*; 36.2 %), plasmatocytes (PM:

165   *Hml*, *Pxn*, *NimC1*; 57.6 %), crystal cells (CC: *lz*, *PPO1*, *PPO2*; 1.3 %), lamellocytes

166   (LM: *atilla*; 1.5 %), the PSC (*Antp*, *col*; 0.9 %), and two additional clusters with

167   uncharacterized genetic features. One novel cluster is enriched with glutathione-

168   S-transferase transcripts including *GstD1*, *GstD3*, *GstE1, GstE7,* and *GstT1*, that we

169   named "GST-rich" (1.0 %). The other novel cluster exhibits high expression

170   levels of phagocytosis receptor-, lipid metabolism-related, and starvation-

171   induced genes such as *crq, eater*, *Sirup*, *LpR2*, and *Lsd-2*. We referred to this

172   cluster as "adipohemocyte" (1.4 %) based on the name of similar hemocytes in

173   other insects (Hillyer et al., 2003). We verified the presence of GST-rich and

174   adipohemocyte cell populations in wild-type lymph glands by confirming the

8

175   expression of signature genes for these clusters in matched bulk RNA-seq

176   (Figure S1H). Additionally, we identified cells of the dorsal vessel (DV; 0.1 %) as

177   an extra non-hemocyte cell type based on previously identified marker genes,

178   *Mlc1* and *Hand*, for this tissue (Figure 1C; Figure S1G).

179   Separation of clustered cells by developmental time points revealed that

180   the relative population size of cell clusters changes as the lymph gland matures.

181   Hemocytes in the lymph gland at 72 h AEL are subdivided into two major

182   groups—prohemocytes and plasmatocytes, with a virtually identical ratio of

183   49.8 % and 46.1 %, respectively (Figure 1D-1E). As the lymph gland matures, the

184   proportion of plasmatocytes exceeds that of prohemocytes, and only 30 % of the

185   hemocytes retain the prohemocyte signature at 120 h AEL (Figure 1D-1E). This

186   result is consistent with proportional changes of prohemocytes or plasmatocytes

187   visualized by marker gene expression during development *in vivo* (Figure S1F).

188   Different from plasmatocytes emerging from 72 h AEL, crystal cells and GST-rich

189   cells first appear at 96 h AEL, and lamellocytes and adipohemocytes appear later

190   at 120 h AEL (Figure 1D-1E). The PSC maintains constant cell numbers and

191   relative ratios across lymph gland development (Figure 1E). Due to temporal

192   discrepancies in the emergence of mature hemocytes, we observed that most

193   cells at 72 h AEL overlap well with cells in subsequent time points, while cells at

194   96 and 120 h AEL segregate from those of preceding points on the *t*-SNE plot

195   (Figure S1I). These results were reproduced by an independent analysis using

196   UMAP and an alternative batch correction method (Figure S1J, see Methods for

197   more details).

198   To better characterize the major cell types and transitions in gene

199   regulatory networks during lymph gland development, we applied SCENIC, an

9

200   algorithm that reconstructs gene regulatory networks and identifies cell states

201   from scRNA-seq data (Aibar et al., 2017). We identified previously recognized

202   transcriptional regulators such as *jumu*, *Antp*, and *kn* (also known as *collier*) in

203   the PSC; *srp* in plasmatocytes; and *lz* and *Su(H)* in crystal cells (Figure S1K)

204   (Crozatier et al., 2004; Hao and Jin, 2017; Lebestky et al., 2003; Mandal et al.,

205   2007). Moreover, we characterized transcription factors in well-known

206   complexes or pathways in each cell type. In prohemocytes, we detected

207   transcription factors of DREAM (Sadasivam and DeCaprio, 2013), a protein

208   complex responsible for cell cycle regulation, including *E2F2* and *Dp,* and the Dpp

209   pathway transcription factor *Mad* at 96 to 120 h AEL (Figure S1K).

210   Plasmatocytes, on the other hand, utilize distinctive transcriptional regulators of

211   the ecdysone pathway highlighted by *br*, *EcR, usp, Eip74EF,* and *Hr4,* and stress

212   responsive genes such as *foxo* and *dl* (Figure S1K). Overall, our single-cell dataset

213   of the entire lymph gland reliably reveals seven major types of hemocytes (PSC,

214   prohemocytes, plasmatocytes, crystal cells, lamellocytes, and two newly

215   identified populations—GST-rich and adipohemocytes). Also, SCENIC analysis

216   delineates transcriptional transitions of the hemocytes and their regulators at

217   the single-cell level.

218

219   **Heterogeneous populations of lymph gland hemocytes**

220   Our scRNA-seq data prompted us to further catalog the heterogeneity of

221   primary cell types by performing unsupervised clustering. With subclustering

222   analysis, we identified eleven subclusters of prohemocytes, ten subclusters of

223   plasmatocytes, and two subclusters each for lamellocytes and crystal cells

224   (Figure 2A; Figure S2A). We ensured that each subcluster contains cells from all

225   libraries except stage-specific subsets (Figure S2B). In addition, we excluded

226   non-hemocyte subclusters enriched with ring gland- or neuron-specific genes

227   (Figure S2C-S2D). Adipohemocytes split into three subordinate clusters; yet, two

228   subclusters were library specific, and thus, only one subcluster was considered

229   relevant and kept for subsequent analyses (Figure S2B). Interestingly, both the

230   PSC and GST-rich clusters did not split into subclusters (Figure S2A).

231   The majority of prohemocyte subclusters is evenly represented at all

232   timepoints and maintains high levels of *Tep4* and *Ance* throughout (Figure 2B).

233   Apart from their constant expression, we identified discrete fluctuations of cell-

234   cycle regulating genes including *polo*, *Cdk1*, *aurB*, *Det*, *CycB*, and *stg* within

235   prohemocytes, accompanied by alterations in additional nuclear genes such as

236   *dUTPase*, *Pen*, and *sle* (Figure 2B). These genes peak at PH1, PH2, and PH4–PH5

237   within prohemocytes, and a comparable pattern is also present in PM3–PM4

238   (Figure 2B). There are obvious distinctions between genes involved in the cell

239   cycle: *stg* and *CycB* are regulators of the G2 phase; *Cdk1* of the G1 phase; and

240   *polo*, *aurB*, and *Det* are regulators of the M phase (Edgar et al., 1994; Llamazares

241   et al., 1991; Mathieu et al., 2013; Parry and O'Farrell, 2001). Based on relative

242   levels of these genes, PH1 is likely to be in G2 and M; PH2 in G2; PH4 in G1; PH5

243   in M; and PM3 and PM4 in M phase (Figure 2B). Similar to prohemocytes,

244   plasmatocytes exhibit gradual changes in *vkg*, *NimC1,* and *eater* while keeping

245   *Hml* and *Pxn* expression high (Figure 2B). PM7 to PM10 express characteristic

246   signature genes such as *Ama*, *vir-1*, and *crq*, only detected at 120 h AEL (Figure

247   2A, Figure S2B). Crystal cells are divided into two groups, CC1 and CC2 (Figure

248   2A). CC1 expresses low levels of *lz* along with the expression of MZ and CZ

249   markers. However, CC2 is devoid of the MZ or CZ markers and only displays high

11

250   levels of *PPO1* and *PPO2*, suggesting that CC1 and CC2 correspond to early and

251   mature crystal cells, respectively (Figure 2A-2B). Similarly, lamellocytes are

252   separated into premature LM1 and mature LM2 reminiscent of the CC1 and CC2

253   clusters (Figure 2A-2B).

254   We next sought to identify new markers and characteristic gene

255   expression patterns in the lymph gland. We confirmed the expression of *Ilp6*, *tau*,

256   *mthl7*, *brat,* and *chrb* in the PSC; *Men* and *Numb* in crystal cells; and *vir-1* and

257   *Ama* in plasmatocytes (Figure 2B, Table S2; Figure S2E-S2G). In addition, we

258   discovered that *zfh1* and *tep2* are expressed in prohemocytes in addition to

259   representative markers such as *Ance* and *dome* in the MZ (Figure S2H-S2I). *crq*,

260   *vir-1*, and *Ama* are significantly expressed in both adipohemocytes and mature

261   plasmatocytes; however, adipohemocytes exhibit high levels of *crq* and *Lsd-2*

262   while keeping low levels of *NimC1* (Figure 2C). In addition to markers widely

263   used (Evans et al., 2014; Yu et al., 2018), we identified new enhancer-trap or

264   MiMIC lines (Nagarkar-Jaiswal et al., 2015) targeting the lymph gland marker

265   genes (Table S3). Lastly, we confirmed the mRNA expression pattern of

266   previously reported genes in each subset (Banerjee et al., 2019) (Figure S3J).

267   Together, we classified 28 transcriptionally distinctive subtypes of hemocytes in

268   the developing lymph gland and assigned functional descriptions of each subset

269   based on gene expression patterns. *Bona fide* markers elucidated in each

270   subcluster collectively provide a valuable resource for further understanding of

271   myeloid hemocyte development.

272

273   **Trajectory reconstitution and functional networks**

274   Hematopoietic events involving transitions of hemocytes from their stem-

275   like to final cell types have been a major focus for understanding *Drosophila*

276   hematopoiesis. Thus, we investigated the time sequence of lymph gland

277   hematopoiesis by reconstruction of developmental trajectories using Monocle 3

278   (Cao et al., 2019). For the trajectory analysis, we excluded PSC as the PSC cells do

279   not give rise to the rest of lymph gland hemocytes (Figure S3A) (Crozatier et al.,

280   2004; Mandal et al., 2007), and we set the PH1 subcluster as the start point

281   based on the expression of *Notch*, *shg*, and high levels of mitotic genes (Figure

282   2B). Pseudotime reconstitution of lymph gland hemocytes displays the main

283   trajectory from prohemocytes to plasmatocytes along with divergent sub-

284   trajectories towards crystal cells, adipohemocytes, GST-rich, and lamellocytes

285   (Figure 3A-3B; Video S1). The trajectory corresponds well with on-and-off

286   patterns of marker genes in the lymph gland (Figure S3B). Moreover, there is an

287   excellent correlation between the real-time and the pseudotime trajectories

288   when compared with segregated real-time hemocyte transcriptomes (Figure 3B-

289   3C, Figure S5C). These analyses validate the *in silico* algorithm-based sequence of

290   hemocyte differentiation and confidently illustrate the developmental phases of

291   lymph gland hemocytes.

292   In the major trajectory, we observed a linear trajectory from PH1 to PH3,

293   projecting towards diverse differentiating states of prohemocytes including PH4-

294   PH8 and GST-rich (Figure 3A, Figure S3D-S3E). In the later sequence, all the PH

295   subclusters including the GST-rich merge into PM1 in the trajectory, implying

296   that GST-rich is a tributary of prohemocytes joining the main PH-to-PM flow

297   (Figure S3D-S3E). A branch is observed following PM1, producing separable

298   paths towards either the plasmatocyte, the crystal cell, or the lamellocyte

13

299   lineages. PM3 is biased to the plasmatocyte and the crystal cell lineages, while

300   PM4 gives rise to late plasmatocytes, adipohemocytes, and lamellocytes (Figure

301   S3D-S3E). We also observed a coupling of cell division and differentiation.

302   Besides PH1 and PH2 subtypes expressing high levels of cell cycle genes, PH4-

303   PH5 and PM3-PM4 emerge immediately after each divergence (Figure 2B; Figure

304   S3D-S3E). As an auxiliary route, PH9 and PH11 are distinguishable from PH3 and

305   bypass the classical PH-to-PM flow to give rise to late plasmatocytes or

306   lamellocytes at 120 h AEL (Figure 3A-3B, Figure S3D-S3E).

307   To address the functional characteristics of the hematopoietic trajectory

308   and associated subclusters, we examined subtrajectory- and subcluster-specific

309   gene-expression modules to determine whether subclusters share similar gene

310   expression modules (Figure S3F). Strikingly, prohemocyte subclusters exhibit

311   related translation-, metabolism-, and signaling gene expression whereas

312   plasmatocyte subclusters show relatively high levels of extracellular matrix

313   (ECM), cytoskeletal and immune responsive genes. Crystal cell subclusters

314   display high levels of genes involved in sugar metabolism, and adipohemocyte

315   and GST-rich subclusters show fatty acid-related and DNA damage responsive

316   gene modules, respectively (Figure S3F).

317   We next focused on the transition of prohemocytes into mature

318   hemocytes and defined subclusters spanning the intermediate zone according to

319   the trajectory analysis and modular configurations. Given that PH5 and

320   PM3/PM4 are mitotic and PH6, PH8, PH10, GST-rich, and PM1, subclusters

321   between PH5 to PM3/PM4, exhibit moderate levels of *Tep4*, *Ance*, *Pxn* and *Hml*,

322   we hypothesized that PH6, PH8, PH10, GST-rich, and PM1 correspond to

323   intermediate cell types prior to differentiation. These subclusters are found at

14

324   all-time points (Figure S2B). We scored highly enriched genes in the potential

325   intermediate subclusters and noticed that expression of *Nplp2* aligns well with

326   these cell populations (Figure 2B). Visualizing *Nplp2* in the lymph gland revealed

327   a partial overlap of *Nplp2* with the MZ marker, *Dome^{Meso}*, or the CZ marker, *Pxn*.

328   However, as *Nplp2* is expressed independently from the late plasmatocyte

329   marker, NimC1 (Figure S3G), it indicates that *Nplp2* is expressed in the

330   intermediate zone of the lymph gland, which corresponds to hemocytes in

331   transition towards differentiation. Altogether, the pseudotime trajectory analysis

332   provides a detailed basis for prohemocyte differentiation. In addition, we

333   demonstrate the presence of subclusters in transition, previously described as

334   the IZ, and their endogenous gene expression.

335

336   **Initiation of hematopoiesis in the lymph gland**

337   Prohemocytes have been established as the precursors of lymph gland

338   hemocytes that produce the entire lymph gland hemocytes (Jung et al., 2005;

339   Minakhina and Steward, 2010b). Despite previous attempts to understand the

340   onset of larval hematopoiesis, it has been unclear whether there is a premature

341   state of prohemocytes reminiscent of mammalian hematopoietic stem cells

342   (HSCs).

343   To investigate the primordial cell types during lymph gland

344   hematopoiesis, we focused on the earliest PHs—PH1 and PH2—which initiate

345   the entire trajectory. We observed that the majority of PH1 is detached from

346   PH2/PH3 and a subset is connected to PH2 in the trajectory map (Figure 2A, 3A,

347   Figure S2A, S3B-S3D). Though PH1 and PH2 mark the earliest pseudotime, both

348   clusters are found at all developmental time points (Figure 3B-3C, Figure S2B).

349   We identified multiple signature genes in the PH1 subcluster (Figure 2B, 3D).

350   First, we discovered that *Notch* (*N*), its ligand, *Delta* (*Dl*), and the *E(Spl)* family

351   genes, downstream targets of the Notch pathway, are expressed in PH1 (Figure

352   3D, Figure S4A). Interestingly, cells in PH1 and PH2 are sequentially arrayed

353   according to on-and-off patterns of *Dl* and *N* (Figure 3E, Figure S4B). Second, we

354   found associations of Hippo, MAPK, Wnt, and Notch pathways with Dl^+N^+ cells of

355   PH1 by KEGG pathway analysis (Figure S4B). Third, we observed levels of *dome*,

356   *hop*, *Stat92E*, and *Socs36E* in PH1, reflecting active JAK/STAT signaling (Figure

357   3D, Figure S4A-S4B). Strikingly, the expression of Notch/Delta and JAK/STAT-

358   related genes in PH1 decrease in the succeeding PH2, suggesting that PH1 cells

359   undergo a drastic change. Fourth, we identified that PH1 does not express *col*

360   while PH2 exhibits low levels of *col* expression, consistent with previous

361   observations (Figure S4C)(Benmimoun et al., 2015). Lastly, high levels of cell

362   cycle genes are detected in both PH1 and PH2, constituting some of the few PH

363   subclusters actively proliferating in the lymph gland (Figure 2B).

364   Next, we applied SCENIC to further establish gene regulatory networks of

365   PH1 and PH2 cell populations with PSC as a comparison. SCENIC analysis on the

366   PSC successfully proved the activity of known transcription factors (Figure 3F).

367   Moreover, the PH1 subcluster displays transcriptional activity of known

368   regulators, such as *sd* and *Stat92E*, as well as novel genes, including *jim*, *Psi*, *bowl*,

369   *esg*, *Tet*, *E(bx)*, and the *E(spl)* family (Figure 3F).

370   We next performed spatial reconstructions for PH1 *in vivo*, and profiled

371   the expression of genes newly identified in the subset. Interestingly,

372   *Stat92E::edGFP* is expressed in the cells neighboring the PSC, which are neither

373    $Tep4^+$ MZ nor $Antp^+$ PSC (Figure 4A, Figure S4D-S4E). Similarly, $Stat92E^+$ cells

16

374   show close contact with *col+* cells without having apparent overlaps (Figure 4B).

375   The number of *Stat92E+* cells increases over development, maintaining relatively

376   constant ratios of these cells (Figure 4C). Furthermore, *Stat92E+* cells are gone

377   upon genetic ablation of the PSC, which indicates that expression of *Stat92E+*

378   PH1 is dependent upon the PSC (Figure 4D). We additionally detected *Dl* mRNA

379   or Dl protein expression near the PSC similar to the pattern observed with

380   *Stat92E::edGFP* (Figure 4E, Figure S4F). The majority of *Dl+* cells are *Stat92E+*;

381   however, Dl covers a range broader than a few cell diameters than *Stat92E+*

382   (Figure 4F). We screened Gal4 lines of signature genes in PH1 and identified a

383   new *Dl* enhancer-trap that marks *Antp-* cells adjoining to the PSC, which produce

384   hemocytes of the entire lymph gland (Figure 4G). To summarize, PH1, the initial

385   subset of the pseudotime trajectory, indicates a novel subpopulation of

386   prohemocytes that physically interacts with the PSC and is adjacent to *col+* PH2.

387   PH1 cells do not co-localize with conventional MZ or CZ markers but exhibit

388   distinctive gene regulatory networks primarily steered by the Delta/Notch and

389   JAK/STAT pathways. Moreover, these cells retain potentials to give rise to

390   hemocytes in the lymph gland during 72 to 120 h AEL (Figure 4H). Thus, we

391   conclude that PH1 is a premature state of prohemocytes reminiscent of

392   mammalian hematopoietic stem cells (HSCs).

393

394   **Differentiation of lymph gland hemocytes upon wasp infestation**

395   We next investigated emerging heterogeneity and differentiation of the

396   lymph gland hemocytes upon wasp infestation. We aligned and matched lymph

397   gland hemocytes from 24 h PI (post-infestation; 96 h AEL) to control hemocytes

398   using the label transfer, that resulted in the annotation of six hemocyte clusters

17

399   (iPSC, iPH, iPM, iCC, iLM, and iGST-rich) and 21 subclusters (9 iPHs, 6 iPMs, 2

400   iCCs, 2 iLMs, iPSC and iGST-rich) when compared to those from controls (Figure

401   5A-5B; Figure S5A). Consistent with previous studies (Crozatier et al., 2004;

402   Ferguson and Martinez-Agosto, 2014), wasp infestation significantly reduces

403   crystal cells (iCCs)—iCC1 and iCC2 (Figure 5A-5B, Figure S5B). A similar decline

404   is readily observed in iPH1, which is confirmed by the reduction of $Stat92E^+$ or

405   Dl^+ iPH1 in the lymph gland upon wasp infestation (Figure 5C, Figure S5B). In

406   contrast, iPH4, iPH6, iPH7, iPM1, and iPM4 show a stark increase in numbers and

407   relative proportions, implying an expansion of differentiating hemocytes upon

408   wasp infestation (Figure S5B). Coinciding with the increase of differentiating

409   cells, the lamellocyte and GST-rich populations, which are barely observed

410   during normal development at 96 h AEL, expand upon wasp infestation (Figure

411   5A-5B, Figure S5B). Lamellocytes derived upon infestation (iLMs) are

412   subclustered into two groups: iLM1 and iLM2, which represent immature and

413   mature iLMs, respectively (Figure S5C). While other cell types undergo

414   significant modifications upon wasp infestation, we did not detect any changes in

415   the expression and number of iPSC upon wasp infestation (Figure 5B, Figure

416   S5D-S5F). When the signature genes of each subcluster are compared to those

417   from controls, gene expression patterns in general are not altered (Figure S5D).

418   However, the intermediate cell population already expresses lamellocyte

419   markers such as *atilla* and *mthl4*, a novel LM marker (Figure S5D). These data

420   indicate that the active immunity causes a biased commitment of prohemocytes

421   and plasmatocytes to the lamellocyte lineage.

422   To better understand how iLM differentiates in the lymph gland, we

423   performed pseudotime trajectory analysis and examined gene modules of

18

424   related subtypes. Upon the trajectory analysis, we discovered that iLMs are

425   associated with a wide span of iPHs and iPMs (Figure 5D, Figure S5F-S5G). The

426   majority of iLMs are directly derived from iPH8 and iPM1 (route 1 and 2 in

427   Figure 5E), subclusters indicated as the intermediate cell populations (Figure 5E,

428   Figure S5H). Additionally, an alternative route is generated from iPM6 (route 3

429   in Figure 5E), the most mature plasmatocyte subcluster at 96 h AEL (Figure 5E,

430   Figure S5H). We validated the data by tracing the IZ or the CZ markers upon

431   wasp infestation and confirmed that L1$^+$ iLMs are derived from either *Nplp2$^+$*

432   intermediate hemocytes or *Hml$^+$* plasmatocytes (Figure 5F). However,

433   differentiating crystal cells and lamellocyte lineages are mutually exclusive

434   (Figure S5I). An association of gene-expression modules of each subcluster

435   indicates the existence of two distinct trajectories to iLMs: iPH8/iPM1-to-iLM

436   and iPM6-to-iLM (Figure 5E, right). The first iPH8/iPM1-to-iLM wave is enriched

437   with genes involved in hemocyte proliferation and oxidative phosphorylation

438   (Figure 5E, right). The second iPM6-to-iLM wave expresses Toll/Imd pathway

439   and structural genes, demonstrating at least two modes of iLM development in

440   the lymph gland upon wasp infestation. Overall, we established that the lymph

441   gland responds to wasp infestation by an expansion of differentiating hemocytes

442   accompanied by subsequent differentiation of iLMs. In addition to the

443   differentiation of intermediate populations indicated as iPH8/iPM1-to-iLM in the

444   trajectory, mature plasmatocytes, iPM6, trans-differentiate into iLMs as an

445   alternative route amplifying the magnitude of iLM formation.

446

447   **Genetic comparison between two hematopoietic lineages**

19

448   *Drosophila* hematopoiesis occurs in two waves, and hemocytes

449   originating from these two lineages differentiate into indistinguishable cell types

450   (Bazzi et al., 2018; Ghosh et al., 2015; Sanchez Bosch et al., 2019). To distinguish

451   and compare these two lineages, we compared the larval circulating hemocyte

452   dataset (see accompanying paper, Tattikota *et al.*) to explore lineage-specific

453   features of *Drosophila* hemocytes at 96 and 120 h AEL. We performed Seurat

454   alignment to cluster datasets after adjusting for batch effect. We also excluded

455   genes related to stress responses that may have been induced during sample

456   preparation (see Methods for details). As a result, we found that hemocytes from

457   the lymph gland significantly overlap with those from circulation (Figure 6A,

458   Figure S6A). We then transferred subcluster labels of lymph gland hemocytes to

459   circulating hemocytes, and recognized three common cell types: prohemocytes,

460   plasmatocytes, and crystal cells, all of which consisted of seven subclusters in

461   circulation (Figure 6B, Figure S6B). Lamellocytes, adipohemocytes, and the PSC

462   are exclusively found in the lymph gland (Figure 6B). All 67 prohemocytes in

463   circulation are labeled as PH1 with unique markers (Figure S6B-S6C), albeit in

464   the absence of Notch and its downstream components (Figure 6C, Figure S5C).

465   Plasmatocytes in circulation share similarities to PM4, PM5, PM6, and PM7 of the

466   lymph gland (Figure S5D). Additionally, crystal cells in the lymph gland and in

467   circulation are nearly identical except for a few genes (Figure S5E-S5F).

468   Next, we explored the collective signature gene expression of lymph gland

469   and circulating hemocytes. Besides the genes depicted caused by uneven

470   proportions (Figure S6A), we identified novel lineage-specific genes including

471   *28SrRNA-Psi:CR45855* and *28SrRNA-Psi:CR45859* in circulation and *CG44250* and

472   *CG33460* in the lymph gland (Figure 6C; Figure S6A). When each subcluster was

20

473   individually compared, circulating plasmatocytes display *Ubx* expression while

474   plasmatocytes in the lymph gland show *Thor* expression (Figure S5D, S5G-S5I).

475   Similar differences are observed in crystal cells: *Pde1c*, *CAH2*, and *Naxd* are

476   higher in circulating crystal cells whereas *Arc2*, *Oscillin*, *aay*, and *fbp* are

477   significant in crystal cells from the lymph gland (Figure S5E-S5F). Taken

478   together, hemocytes generated from the two independent lineages appear to be

479   predominantly similar; however, they are sufficiently genetically distinct that we

480   can distinguish their ancestries.

481

Evolutionary conservation of lymph gland hemocytes

483   Although functional homologies between *Drosophila* hemocytes and

484   vertebrate immune cells have been addressed previously (Cooper, 1976; Evans

485   et al., 2003; Gold and Bruckner, 2014), no system-level comparison has been

486   reported. Thus, we compared the single-cell transcriptome profiles of six

487   hematopoietic *Drosophila* cell types including prohemocytes, plasmatocytes,

488   crystal cells, lamellocytes, GST-rich, and adipohemocytes, with human

489   hematopoietic lineages from the Human Cell Atlas (HCA) project (Census of

490   Immune Cells)(Rozenblatt-Rosen et al., 2017). 262,638 high-quality immune

491   cells were clustered and annotated with 19 well-known immune cell types

492   (Figure 7A-7B). Following clustering, we observed a clear separation of

493   lymphoid cells, including T cells and B cells, from the others, while erythroblasts,

494   B-cell precursors, and granulocyte progenitors are closely linked to

495   hematopoietic stem cells~multipotent progenitor (HSC~MPP) cluster (Figure

496   7A-7B).

497   To compare the similarities of gene expression patterns between human

498   immune cells and *Drosophila* hemocytes, 6,463 orthologous gene pairs were

499   retrieved, and enrichment of the top 30 signature genes of each hemocyte type

500   was weighed in human cell types using gene set variance analysis (GSVA) (see

501   Methods for details). Strikingly, we observed a clear separation into two

502   lineages, with one cluster consisting of lymphoid-lineage human cells devoid of

503   *Drosophila* hemocytes, whereas the second one was enriched with myeloid

504   lineage containing most of the hemocytes (Figure 7C; Table S4). Prohemocytes in

505   the *Drosophila* lymph gland share similar gene expression with the HSC~MPP

506   cluster in humans, while mature hemocytes in general are comparable to human

507   myeloid immune cells (Figure 7C). Specifically, plasmatocytes exhibit similar

508   expression with monocytes, dendritic cells, and platelets, whereas

509   adipohemocytes resemble more plasma cells (Figure 7C). Both crystal cells and

510   GST-rich are closely associated with CD14$^+$ monocytes and classical dendritic

511   cells (cDC); however, crystal cells are much closer to granulocyte progenitors

512   than other cell types (Figure 7C). Intriguingly, genetic similarities of *Drosophila*

513   hemocytes and human myeloid cells are extensively affected when immune-

514   challenged hemocytes and human immune cells are compared. Plasmatocytes

515   become comparable to HSC~MPP, and, surprisingly, *Drosophila* hemocytes at 24

516   h post-infestation display striking similarities with lymphoid precursor lineages

517   such as naïve T cells, T cell precursors, and pro-B cells of humans, implying the

518   functional duality of active hemocytes beyond their homologies to myeloid cells

519   (Figure 7D). In conclusion, our analysis demonstrates that *Drosophila* hemocytes

520   retain conserved genetic characteristics of broad classes of human myeloid

22

521   immune cells and are plastic enough to show lymphoid-like features upon active

522   immunity.

23

523   **DISCUSSION**

524   In this study, we report a comprehensive single-cell transcriptome

525   analysis of 29,618 developing myeloid hemocytes in *Drosophila* lymph glands.

526   Our analysis provides insights into: 1) the development of myeloid hemocytes at

527   the single-cell level, 2) the existence of hematopoietic stem-like populations and

528   adipohemocytes in invertebrates, 3) the differentiation mechanisms of myeloid

529   hemocytes upon active immunity, 4) the genetic difference of hemocytes derived

530   from two different hematopoietic ancestries: embryo and larva, and 5) the

531   evolutionary relevance of *Drosophila* hemocytes. To our knowledge, this study is

532   the first description of invertebrate myeloid cells at a single-cell level and a

533   system-level comparison of myeloid cells across species.

534

535   **scRNA-seq reveals novel hemocyte types in the lymph gland**

536   Unlike human hematopoietic systems, *Drosophila* hematopoiesis takes

537   place in a unique hematopoietic organ, the lymph gland, where all developing

538   myeloid cells are located until histolysis upon puparation (Grigorian et al.,

539   2011). Thus, scRNA-seq of the lymph gland allows us to analyze inclusive

540   profiles of myeloid hemocytes containing all developmental stages of myeloid

541   cells. Our scRNA-seq datasets faithfully display single-cell transcriptomes of all

542   known cell types as well as two novel cell types: GST-rich and adipohemocytes.

543   GST-rich cells, enriched with ROS-responsive and DNA damage genes, emerge

544   during prohemocyte development. Considering that genes enriched in GST-rich

545   cells are also evident in the lymph gland bulk RNA-seq, this novel population is

546   not a consequence of stressed hemocytes. Rather, this subtype may represent a

547   state that prohemocytes experience during development, or may play an active

548   role in ROS- or GABA-mediated stress responses (Madhwal et al., 2019; Owusu-

549   Ansah and Banerjee, 2009; Shim et al., 2013). Adipohemocytes, on the other

550   hand, share hallmarks of both mature plasmatocytes and lipid metabolism,

551   appearing only at 120 h AEL of the lymph gland. Macrophages in vertebrates

552   readily take up lipids and lipoproteins, and accumulation of lipid-containing

553   macrophages, called foam cells, is highlighted in various pathological conditions

554   (Li and Glass, 2002; Moore et al., 2013). In *Drosophila*, the presence of lipid-

555   containing hemocytes has not been reported. Given our analyses, and that

556   adipohemocytes are frequently observed in insects, including *Aedes aegypti*

557   (Hillyer et al., 2003), it is possible that flies also conserve metabolism-oriented

558   hemocytes to coordinate immunity and metabolism.

559

560   **Prohemocytes are comprised of heterogeneous cell types and states**

561   Prohemocytes have been widely considered to represent a uniform cell

562   population based on the expression of marker genes, *domeless* or *Tep4*. However,

563   recent studies have suggested that prohemocytes may be more heterogenous

564   based on uneven expressions of cell cycle markers or bifurcated *col* expressions

565   (Baldeosingh et al., 2018; Sharma et al., 2019). In support of these studies, our

566   unbiased subclustering of primary clusters identified different status of

567   prohemocytes. First, prohemocytes differ in the expression of cell cycle

568   regulators, implying an asynchrony of prohemocyte development and their

569   states. This observation also accounts for the stochastic cell cycle patterns

570   visualized with the UAS-FUCCI system, a fluorescent-based cell cycle indicator

571   (Sharma et al., 2019; Zielke and Edgar, 2015). Second, we observed dynamic

572   expression patterns of immunogenic, metabolic or stress-responsive genes in PH

573   subclusters. For example, PH7 and PH10 are immunogenic while PH4, PH5, PH6,

574   PH8, and GST-rich are metabolic or stress-responsive. These multiple states

575   could be due to different susceptibility of prohemocytes to physiological

576   conditions and may directly influence lineage specification. Lastly, the presence

577   of prohemocytes with more differentiated states is also indicative of their

578   dynamics. Although the presence of the intermediate zone has been recognized

579   in previous studies (Krzemien et al., 2010; Owusu-Ansah and Banerjee, 2009),

580   the biological significance of various intermediary states and the novel functions

581   of expressed endogenous genes including *Nplp2* in these subclusters have not

582   been explored.

583

584   **PH1 is the stem-like prohemocyte in the lymph gland**

585   As the most primitive subcluster identified in this study, PH1, demarcates

586   a group of cells that has not been annotated by previous markers such as *Tep4*,

587   *Antp* or *col*. Discovery of the hidden cell population – PH1, will shed light on

588   understanding the hierarchy of prohemocyte differentiation and enhance the

589   relevance of the lymph gland as a hematopoietic model. Roles for *Notch*, *Stat92E*,

590   or *scalloped* in the lymph gland development have been previously suggested by

591   recent studies (Dey et al., 2016; Ferguson and Martinez-Agosto, 2017; Krzemien

592   et al., 2007; Mondal et al., 2011). Moreover, clonal analyses have shown that cells

593   adjacent to the PSC generate the largest population in the lymph gland

594   (Minakhina and Steward, 2010a). These studies are consistent with our

595   hypothesis that Notch/Delta- and JAK/STAT-positive cells nearby the PSC

596   sustain latent capacities to produce the entire lymph gland hemocytes. We also

597   observed that PSC is required for the expression of PH1, and thus, PSC is likely to

26

598   provide necessary signals for the maintenance of PH1 during normal

599   development. Given that *col*⁺ hemocytes, referred as PH2 in this study, are

600   independent of the PSC (Baldeosingh et al., 2018; Benmimoun et al., 2015), it is

601   possible that there are multiple niches for respective subtypes, such as the PSC

602   for PH1 and the dorsal vessel for PH2/PH3, which are reminiscent of context-

603   dependent niches found in vertebrates (Tikhonova et al., 2019). Identification of

604   the factors underlying the maintenance and differentiation of PH1 will be

605   important for understanding the nature of hematopoietic stem-like populations.

606

607   **Hemocytes conserve genetic homologies with human immune cells**

608   We attempted to uncover cellular heterogeneity of identical cell types

609   originating from two different lineages and genetic homologies between

610   *Drosophila* hemocytes and human immune cells. Previously, *Drosophila*

611   hemocytes have been proposed to be most akin to macrophages of vertebrates

612   (Franc et al., 1996; Sanchez Bosch et al., 2019); however, our analysis indicates

613   that *Drosophila* hemocytes show characteristics of multiple human myeloid cells,

614   including monocytes, dendritic cells, and granulocyte progenitors. Furthermore,

615   our analysis reveals that *Drosophila* hemocytes additionally acquire signatures of

616   lymphoid lineages upon active immunity. In light of our analyses, we propose

617   that invertebrate myeloid cells carry hidden elements of lymphoid activity,

618   contributing to divergence of the lymphoid lineage in vertebrates. Collectively,

619   our comparative analysis supports genetic similarities between myeloid cells of

620   flies and vertebrates, providing a resource to further understand the

621   invertebrate and vertebrate myeloid cells.

622

27

Acknowledgements

Author Contributions

B.C., S.G.T., D.L., F.K., N.C., M.S., H.D., and S.Y.O. performed experiments; B.C., S.Y., S.G.T., Y.H., J.N., and J.S. analyzed data; S.Y.O. and S.J.M. provided technical support for Drop-seq; B.C., S.Y., F.K., S.G.T., N.P., J.N., and J.S. contributed to writing the manuscript; N.P., J.N., and J.S. supervised the project; J.N., and J.S. conceived the idea.

Declaration of Interest

The authors declare no competing interests.

645 **Figure Legends**

646 **Figure 1. Major cell types identified in developing *Drosophila* lymph glands**

647 (A) *Drosophila* lymph glands (blue, DAPI) at three timepoints (72, 96, and 120 h

648 AEL; After Egg Laying) (left). Schematic workflow of sample preparation for

649 scRNA-seq using Drop-seq (right). Scale bar, 30 μm. Lymph glands are

650 demarcated by white dotted lines.

651 (B) DAPI-positive cell counts of a single lymph gland lobe (n = 30 each for three

652 time points). Red horizontal lines show median counts (397, 1392, and 4557 for

653 72, 96, and 120 h AEL, respectively).

654 (C) A t-SNE plot showing the two-dimensional projection of eight major cell

655 types identified in the scRNA-seq dataset (n = 22,645). The count of each cell

656 type is indicated in parentheses. Colors denote cell types. Dotted lines demarcate

657 prohemocytes (blue) and plasmatocytes (red).

658 (D) Two-dimensional projections of major cell types along developmental time

659 points (left) and proportion of the cell types at each time point (right).

660 Proportions of prohemocytes (blue) and plasmatocytes (red) are indicated.

661 (E) Relative proportion (indicated as proportional ratio; top) and normalized cell

662 counts (bottom) of each major cell type. Colors represent sampling time points.

663

664 **Figure 2. Heterogeneous cellular states of hemocytes in the lymph gland**

665 **defined by subclustering analysis**

29

666   (A) Subclusters of hemocytes—prohemocytes, plasmatocytes, lamellocyte, and

667   crystal cells—are projected onto two-dimensional *t*-SNE plots. Non-

668   hematopoietic cell types (Neurons; RG, ring gland) are indicated. The numbers in

669   the plots represent the subcluster number.

670   (B) Dot plot presentation of significant gene sets in the 31 subclusters. 5

671   representative markers, *srp*, *Tep4*, *Ance*, *Hml*, and *Pxn* are indicated to the left

672   column. Cell-cycle regulating genes are shown in the middle. Signature genes

673   identified in this study are marked with subcluster markers. Dot color shows

674   levels of average expression, and dot size represents the percentage of cells

675   expressing the corresponding marker genes in each subcluster.

676   (C) *Ama* (magenta; *Ama-Gal4 UAS-Red stinger*), *crq* (magenta; *crq-Gal4 UAS-Red*

677   *stinger*) or *Lsd-2* (magenta; *Lsd2-Gal4 UAS-Red stinger*) partially overlap with

678   NimC1 (white). Ama^+NimC1^+, crq^+NimC1^+ and $Lsd\text{-}2^+NimC1^+$ cells correspond to

679   plasmatocytes, whereas Ama^+NimC1^-, crq^+NimC1^- and $Lsd\text{-}2^+NimC1^-$ cells

680   represent putative adipohemocytes. Generally, Ama^+NimC1^-, crq^+NimC1^- and *Lsd-*

681   2^+NimC1^- cells show higher *Ama*, *crq*, or *Lsd-2* expression levels than those of

682   double-positive cells. Magnified images show colocalization of *Ama*, *crq* or *Lsd2*

683   with NimC1 (right in each panel). White scale bar indicates 30 μm; yellow scale

684   bar, 3 μm. White dotted line demarcates the lymph gland.

685

686   **Figure 3. Reconstruction of lymph gland hematopoiesis using pseudotime**

687   **trajectory analysis**

688   (A) A three-dimensional landscape of the lymph gland hematopoiesis trajectory

689   using Monocle 3 (n = 19,143). Non-hematopoietic cells were excluded in this

690   analysis. Black line indicates the trajectory. Colors indicate the six major cell

691   types used for analysis. The inset shows the three ancestral PH subclusters, PH1,

692   PH2, and PH3.

693   (B) Trajectories re-drawn by developmental time points (top) and calculated

694   pseudotime (bottom). Colors indicate the three real-time points (top) and

695   pseudotime (bottom).

696   (C) Relative densities of hemocytes segregated by three time points (top) and

697   cell types (bottom) along pseudotime. PH1 and PH2 are separated from other PH

698   subclusters for higher resolution. Colors in density plots correspond to

699   pseudotime, as in B.

700   (D) Heatmap representation of the 35 signature genes identified in PH1, PH2,

701   and the PSC (n = 77, 79, and 189 cells, respectively). The colored legend denotes

702   the standardized level of the genes.

703   (E) Four subgroups in PH1 and PH2 defined by the expression of *Delta* (*Dl*) and

704   *Notch* (*N*). Colors show subgroups and shapes specify PH subclusters. X axis

705   means *N* expression; Y axis, *Dl* expression.

706   (F) Binary heatmap showing the activity of transcription factors in PH1, PH2, and

707   the PSC predicted by SCENIC. Numbers in parentheses denote the count of

708   downstream genes used to test the activity of transcription factors.

709

710   **Figure 4. Expression of PH1 in the lymph gland**

31

711 (A) *STAT92E+* (green) and *TepIV+* (magenta) or *Antp+* (white) cells are mutually

712 exclusive (*TepIV-Gal4 UAS-mCherry; STAT92E::edGFP*). The dotted box indicates

713 the region magnified. High magnification of *STAT92E+* and *TepIV+* cells near *Antp+*

714 PSC.

715 (B) *STAT92E+* cells (green) do not co-localize with cells expressing high (PSC) or

716 low (PHs) levels of *collier* (magenta). Box indicates magnified view. High

717 magnification of *STAT92E+* and *col+* cells near the PSC.

718 (C) The number of *STAT92E+* cells (green) increases during lymph gland

719 development (72, 96, and 120 h AEL). Exclusive expression of *Antp+* (magenta)

720 and *STAT92E+* (green) is maintained at all time points. Graphs represent

721 quantitation of the number (left) or the proportion (right) of *STAT92E+* cells in

722 one lymph gland lobe.

723 (D) Genetic ablation of the PSC (*pCol85-Gal4; STAT92E::edGFP UAS-hid, rpr*)

724 attenuates *STAT92E* (green) expression in the lymph gland (left). Graph indicates

725 quantitations of the number of *STAT92E+* cells in one lymph gland lobe (right,

726 ***$P<0.0001$).

727 (E) Dl+ cells (magenta) are localized adjacent to the PSC (Antp, green). Box

728 indicates the magnified area. Magnified views of Dl+ (magenta) and *Antp+* (green)

729 cells (two markers, right top; one marker, right bottom). Cyan dotted lines

730 delineate Dl+ expressing cells.

731 (F) Dl+ cells (magenta) co-localize with *STAT92E+* (green). Box indicates the

732 magnified area. Magnified view of Dl+ (magenta) and *STAT92E+* (green) cells

733   (right). A few Dl⁺ cells that do not express *STAT92E::edGFP* are indicated

734   (arrowhead).

735   (G) Lineage tracing of *Dl⁺* cells (green, traced; magenta, real time; blue, Antp).

736   *Delta-Gal4 UAS-GTRACE* covers the entire lymph gland. Box indicates the

737   magnified view. Arrowheads represent *Dl⁺* cells next to the *Antp⁺ Dl⁺* PSC. The

738   PSC does not give rise to lymph gland hemocytes (see Supplememtary Figure

739   3A).

740   (H) Model. PH1 cells are adjacent to the PSC. PH1 and PSC or PH2 are mutually

741   exclusive. There are multiple states of prohemocytes including GST-rich and

742   intermediary PHs/PMs. Plasmatocytes represent an heterogenous cell

743   population including adipohemocytes. Lamellocytes are rarely observed under

744   normal conditions. Crystal cells are found among differentiated plasmatocytes.

745   In panels A through G, white scale bar indicates 30 μm; yellow scale bar, 3 μm.

746   White dotted line demarcates the lymph gland. Median value is represented in

747   graphs (C, D).

748

749   **Figure 5. Lymph gland hematopoiesis following wasp infestation**

750   (A) UMAP projections of major cell types defined in normal (top, prefix with 'n')

751   and wasp infested (bottom, prefix with 'i') lymph glands at 96 h AEL (thus, 24 h

752   post infestation (PI)). Non-hematopoietic cell types are excluded. Different colors

753   indicate each cell type.

33

754 (B) Relative proportion (top, proportional ratio) or normalized cell counts

755 (bottom) of major cell types in normal (blue) and wasp infested (red) lymph

756 glands.

757 (C) Wasp infestation reduces *STAT92E+* (left) or Dl+ (right) PH1 populations.

758 *STAT92E* (green) or Dl (magenta) are expressed near the PSC of the lymph gland

759 (top). The expression is attenuated upon immune challenges (middle). Graphs

760 represent quantitation of the number of *STAT92E+* cells (bottom, left) or the

761 proportion of Dl+ (bottom, right) cells in one lymph gland lobe after infestation

762 (24 h PI). Median value is shown in graphs (bottom, ***p<0.0001).

763 (D) A three-dimensional trajectory landscape of major cell types under wasp

764 infestation (left), and additional representation of trajectory over calculated

765 pseudotime using Monocle 3 (right). Box indicates the cells used for

766 subtrajectory analysis in (E). Colors in legends show pseudotime (right).

767 (E) Subtrajectory analysis of five subclusters—iPH8, iPM1, iPM6, iLM1, and

768 iLM2—detected in the trajectory to iLM (left). Two different waves, arrow 1/2

769 (iPH8/iPM1, thus, intermediate cells) and arrow 3 (iPM6), advance towards iLM

770 with distinct gene modules (right). Shared gene modules between iPH or iPM

771 with iLM are indicated in boxes. Colored expression represents the *z*-

772 transformed enrichment level of gene modules.

773 (F) Lamellocytes differentiate from intermediate iPHs (*Nplp2-Gal4 UAS-GTRACE*)

774 or iPMs (*Hml-Gal4 UAS-GTRACE*) upon wasp infestation. *Nplp2+* iPHs (green,

775 traced; blue, DAPI; top) or *Hml+* iPMs (green, traced; blue, DAPI; bottom) express

776 L1 (magenta) upon wasp infestation. Insets indicate magnified images of L1+

777 cells. Cyan dotted lines within insets demarcate traced iLMs.

778   Lymph glands are demarcated by white dotted lines. White scale bar is 30 μm;

779   yellow scale bar is 3 μm.

780

781   **Figure 6. Transcriptome-wide comparisons between embryonic and**

782   **definitive hemocytes in *Drosophila***

783   (A) Two-dimensional projections of hemocytes in the lymph gland (top) and

784   circulation (bottom) at 96 and 120 h AEL.

785   (B) Combined projection (top) and proportions of major cell types (bottom) in

786   the lymph gland (yellow) and in circulation (cyan). Inset (top) shows the ratio of

787   major cell types between the lymph gland (L) and in circulation (C).

788   (C) Dot plot of marker genes highly enriched in a lineage-specific or cell type-

789   specific manner. The colors show the origin of the datasets (yellow, lymph gland;

790   cyan, circulation).

791

792   **Figure 7. Transcriptome-wide comparisons of *Drosophila* hemocytes and**

793   **human immune cells.**

794   (A) Expression of known markers for human hematopoietic cells in the bone

795   marrow. The colored bar indicates the level of scaled gene expression. CD34

796   indicates hematopoietic stem cell; HBA1, erythrocyte; CD79A, B cell; CD3D, naïve

797   T cell; CD8A, CD8 T cell; NKG7, NK cell; LYZ, granulocyte progenitor; CD14,

798   CD14$^+$ monocyte; and FCGR3A, CD16$^+$ monocyte.

799   (B) Annotations of 19 human hematopoietic cell clusters of 262,638 cells

800   provided by the Human Cell Atlas project.

801   (C) Gene set variance analysis (GSVA) of *Drosophila* lymph gland hemocyte and

802   human hematopoietic cell clusters. The colored bar indicates the Gene set

803   variance analysis (GSVA) score.

804   (D) GSVA of hemocytes with human hematopoietic cells from normal (red) and

805   24 h post wasp infestation (blue) lymph glands at 96 h AEL. The colored bar

806   indicates the GSVA score.

807 **Material and method**

808 *Drosophila* **stocks and genetics**

809 The following *Drosophila* stocks were used in this study: *Dome^Meso-EBFP2*

810 (U.Banerjee), *HmlΔ-Gal4* (S.Sinenko), *Dome^Meso-Gal4* (M.Zeidler), *TepIV-Gal4* (NIG,

811 Japan), *Nplp2-Gal4* (KDRC, South Korea), *pCol85-Gal4* (M.Crozatier), *Delta-Gal4*

812 (Bloomington), *Antp-Gal4* (U.Banerjee), *Stat92E::edGFP* (N.Perrimon), *Hml-DsRed*

813 (K.Brueckner), *vir-1^MiMiC* (Bloomington), *Ance^MiMiC* (Bloomington), *UAS-hid, rpr*

814 (Nambu JR), *UAS-GTRACE* (C.Evans), *nSyb-Gal4* (Bloomington), *Ama-Gal4* (NIG),

815 *crq-Gal4* (Bloomington), *Lsd2-Gal4* (NIG), *Ilp6-Gal4* (A.Brand), *Tau^MiMiC*

816 *(*Bloomington*)*, *mthl7-Gal4* (generated in this study), *Chrb^MiMiC* (Bloomington),

817 *Men-Gal4 (*NIG, 113708*)*, *NUMB::GFP* (F.Schweisguth), *zfh1-Gal4 (*Bloomington),

818 *Tep2^MiMiC*(Bloomington), *UAS-EGFP* (Bloomington), *UAS-mCD8GFP*

819 (Bloomington), *Ubx RNAi*(v37823), *lz-gal4^DBD ; Pxn-Gal4^AD* (generated in this

820 study)

821 Fly stocks used in this study were maintained at 25 °C. *Oregon R* was

822 used for the scRNA-seq as a wild type. Unless indicated, crossed flies were

823 maintained at 25 °C with dextrose-cornmeal based normal food.

824 To synchronize larval stages, one hundred adult flies were kept on grape-

825 juice agar plate for two hours. Hatched larvae were discarded at 23 hours after

826 egg laying (AEL), and those at 24 hours AEL were collected and reared on normal

827 corn-meal yeast media. To screen the Gal4 lines in this study, we crossed each

828 Gal4 strain with *UAS-GTRACE* and identified those expressed in the lymph gland.

829 To avoid stress conditions generated by crowding, we reared less than 50 larvae

830 in one vial.

37

831

Generation of fly stocks

To generate Gal4 fly lines, fly genomic DNA was amplified by primers indicated in Supplementary table 3. Amplified genomic regions were ligated into *pAGal4+*(KDRC) or TOPO-TA vector (Invitrogen; K250020) for Gateway cloning. *pBPnlsLexAp65Uw* (Addgene 26230), *pBPZpGAL4DBDUw* (Addgene; 26233) or *pBPp65ADZpUw* (Addgene; 26234) was used as destination vector. Transgenic flies were generated by KDRC, South Korea.

839

Dissociation of lymph glands into single cells

100 to 150 lymph glands were dissected at 72, 96, or 120 h AEL respectively, in Schneider's medium (Gibco, 21720024). Dorsal vessel, ring gland and posterior lobes were detached from the primary lobe of the lymph gland; only the primary lobes of the lymph glands were used in this study. Primary lobes were kept in 200 μl ice-cooled Schneider's medium during dissection. After, centrifugation at 3,000 rpm and 4 ℃ for 5 minutes was done. Supernatant was discarded and 300 μl of room temperature Schneider's medium was added to the lymph gland primary lobes. 300 μl of Papain (Worthington, LK003178) pre-heated to 37 ℃, and 4.1 μl of Liberase TM (Roche, 5401119001) were added and gently mixed. Samples were incubated for 20 minutes with gentle agitation. At 5-, 10-, and 15-minute time points of incubation, samples were mixed using 200p pipette. Enzymes were inactivated with 100 μl of ice-cooled Schneider's medium, and samples were kept on ice. Suspended cells were passed through a 40 μm cell strainer (Corning, 352340). Afterwards, centrifugation at 3,000 rpm and 4 ℃ for 5 minutes was done. The supernatant was discarded, and 1X filtered

38

856   sterile PBS was added to cells. The final concentration of cells was fixed to 300

857   cells/µl for a total of 600 µl.

858

859   **Drop-seq and scRNA-seq**

860   All the Drop-seq and cDNA synthesis methods followed a previous study

861   (Macosko et al., 2015). The concentration of beads was fixed to 300 beads/µl.

862   Around 10 minutes Drop-seq run was performed for each experiment. After

863   cDNA synthesis, scRNA-seq was performed using Illumina NextSeq.

864

865   **Preprocessing and mapping of scRNA-seq data**

866   Raw scRNA-seq data were generated in paired-end reads following

867   single-cell capture using Drop-seq: one end included a barcode and unique

868   molecular identifier (UMI) sequences in 20 nucleotides (12 and 8 nts,

869   respectively), and the other end, cDNA in 50 nts. The preprocessing and mapping

870   of scRNA-seq data produced in this study followed the Drop-seq Core

871   Computational Protocol version 1.2 (January 2016) and corresponding Drop-seq

872   tools version 1.13 (December 2017) provided by the McCarroll Lab

873   (http://mccarrolllab.org/dropseq/).

874   First, the reference genome (Fasta, BDGP 6.02) and transcriptome

875   annotations (gtf, September 2014) required for the processing were downloaded

876   from the Ensembl website (http://asia.ensembl.org/). Additional dictionary and

877   refFlat files were generated using *picard* (*CreateSequenceDictionary*) and

878   *ConvertToRefFlat* provided in the Drop-seq tools package, respectively. These

879   reference data were prepared with the same prefix and stored in a single

880   directory for later use. Simplified command lines are as follows:

881   Generation of a dictionary file:

882 ```
java -jar <path to Drop-seq tools picard>/picard.jar
```

883 ```
CreateSequenceDictionary R=<genome fasta> O=<output dictionary>
```

884   Generation of a refFlat file:

885 ```
<path to Drop-seq tools>/ConvertToRefFlat \
```

886 ```
ANNOTATIONS_FILE=<gft annotation> \
```

887 ```
SEQUENCE_DICTIONARY=<dictionary file> O=<output refFlat>
```

888

889        Once all the reference data was prepared, paired-end fastq files were

890    converted to the bam format using *picard FastqToSam*.

891 ```
java -jar <path to Drop-seq tools picard>/picard.jar FastqToSam \
```

892 ```
F1=<fastq 1> F2=<fastq 2> O=<output bam> SM=<library number>
```

893

894        The unaligned bam files were subjected to the *Drop-seq_alignment.sh*

895    script for alignment to genome. This shell script is a single pipeline that executes

896    detection of barcode and UMI sequences, filtration and trimming of low-quality

897    bases and adaptors or poly-A tails, and alignment of reads using *STAR* (2.5.3a).

898 ```
<path to Drop-seq tools>/Drop-seq_alignment.sh \
```

899 ```
-g <path to STAR index> -r <genome fasta> -n <# of cells expected> \
```

900 ```
-d <path to Drop-seq tools> -s <path to STAR> \
```

901 ```
-o <path to output> -t <path to temporary output> -p <unaligned bam
```
902 ```
file>
```

903

904   **Selection of cells by the total mapped reads**

905   To extract the number of cells having proper read counts, the aligned

906   bam files generated from the previous section were summarized using

40

907 *BAMTagHistogram* in the Drop-seq tools package. This program extracts the

908 number of aligned reads per cell barcode which is subsequently used to plot the

909 cumulative distribution of reads.

910 `<path to Drop-seq tools>/BAMTagHistogram I=<aligned bam> O=<output file>`
911 `                                  TAG=XC`

912

913 Cumulative read distribution plots were then explored, and the number of

914 cells were inferred where a sharp decrease (referred as 'knee' by the author's

915 documentation) in a slope occurs. The inferred cell number was determined as a

916 minimal threshold number of aligned reads per cell for cell selection. To

917 summarize, a minimum of 30,000 reads per cell for 72 h AEL library 6, 15,000

918 per cell for 72 h AEL library 3, 10,000 per cell for 72 h AEL libraries 1, 2, and 4;

919 5000 per cell for 96 h AEL libraries 1, 3, 5, all 120 h AEL libraries, and infested 96

920 h AEL libraries 1, 2, and 4; 4000 per cell for infested 96 h AEL 2, 3000 per cell for

921 96 h AEL library 4 and infested 96 h AEL library 3; 2000 per cell for 92 h AEL

922 library 2 were chosen as thresholds. *DigitalExpression* provides UMI count

923 matrix (selected cells by genes) using a mapped bam file and the minimum

924 number of reads per cell as following.

925 `                <path to Drop-seq tools>/DigitalExpression \`
926 `    I=<aligned bam> MIN_NUM_READS_PER_CELL=<read count threshold> \`
927 `        O=<output read count matrix> SUMMARY=<output summary>`
928

929 The resulting output per library is written into a file with a name where a

930 corresponding library number was added as a suffix to each barcode sequence

931 with an AEL timepoint (*e.g.* barcode-72-1 or barcode-96-2) to avoid collision of

932 barcode sequences between libraries. The number of expressed genes between

933 libraries or sampling timepoints may vary because each library would have a

41

934   different number of captured cells, different cell types, or uneven sequencing

935   depth. In total, at least 13,612, 13,523, 14,277, and 13,658 genes, and 2505,

936   10,027, 11,702, and 10,939 cells were detected in one library at 72, 96, 120, or

937   infested 96 h AEL respectively. So, we used a union set of genes (15,540 genes)

938   to merge three normal lymph gland datasets.

939

940   **scRNA-seq data analysis using Seurat 3.0**

941   Seurat is a universal software for scRNA-seq analyses including

942   preprocessing, cell clustering, and dimension reductions. The current version

943   (v3.0) of Seurat features dimension reduction using uniform manifold

944   approximation and projection (UMAP), integration of datasets produced with

945   different modalities or conditions, and transfer of cell labels between datasets

946   (Butler et al., 2018; Stuart et al., 2019). Detailed analyses steps are explained on

947   the Seurat website (https://satijalab.org/seurat/), so we only describe the

948   schematic workflow used in this study.

949   First, each library was filtered for low-quality cells, separately, by setting

950   thresholds for UMI and gene counts. We used 5,000 genes as an upper threshold

951   and 400 genes as a lower threshold for normal lymph gland libraries and

952   infested lymph gland library 1 and 2, and 200 genes as a lower threshold for

953   other infested lymph gland libraries. Then we also filtered cells having UMI

954   counts higher than two standard deviations from the mean UMI count to exclude

955   multiplets. After filtration, 2399, 9496, 11,081, and 10,461 cells remained for 72,

956   96, 120, and infested 96 h AEL, respectively. All libraries in each sampling

957   timepoint (h AEL) were then merged and normalized, and cells expressing

42

958 mitochondrial genes higher than 10% of total UMI count were removed (2322,

959 9411, 10,976, and 10,179 cells remained). After high mitochondrial cells were

960 excluded, multiplets were inferred using Scrublet (Wolock et al., 2019), which

961 simulates artificial doublets using given expression matrix to predict multiplet

962 artifacts. One, 11, and 52 cells were further filtered out from 72, 96, and 120 h

963 AEL lymph glands. However, no cells were detected from the infested dataset. As

964 a result, 2321 (72 h AEL), 9400 (96 h AEL), 10,924 (120 h AEL), and 10,179 cells

965 (infested 96 h AEL) were subjected to downstream analyses.

966 Next, for the integration of cells from normal lymph glands at three

967 timepoints (normal lymph gland integration), cells were aligned using

968 *FindIntegrationAnchors()* and *IntegrateData()* with default parameters,

969 respectively. UMI counts were normalized, log-transformed, and scaled to

970 properly integrate datasets, and 52 principle components (PCs) were used to

971 explain the variability of the scaled UMI counts across cells. t-SNE and UMAP

972 plots were then manually curated with random seeds using the selected PCs.

973 Clustering was performed with resolution of 0.8 to get 19 clusters. Then again,

974 clusters were aggregated to get broad cell types based on expression of known

975 marker genes (Figure S1G). In summary, six and seven clusters were merged as

976 collective PH and PM, respectively, to define the following six major cell types:

977 PSC, LM, CC, DV, GST-rich, and adipohemocyte.

978 When the integrated normal lymph gland dataset was examined, we

979 found that subclusters of PH and PM cells solely originated from 120 h AEL

980 (Figure 1D and Figure S1I). Thus, we analyzed our dataset using a different batch

981 normalization method to test whether this trend is independent of our analysis

43

982 strategy. For this, we corrected the sequencing library variable using *ScaleData*

983 function along with the UMI count. We then selected the number of PCs to use

984 (50 PCs) and performed t-SNE and UMAP analyses again. Similarly, a number of

985 cells from 120 h AEL were separated from others while cell types, such as PSC,

986 LM, or CC were mixed together (Figure S1J). An R package, *rgl*, was used for all 3-

987 dimensional plots presented in this study.

988

989 **Pseudo-bulk RNA-seq trajectory analysis**

990 Normal lymph gland scRNA-seq libraries were examined to see whether

991 they could be aligned in a single trajectory line as actual development timepoints.

992 As we produced at least four sequencing libraries for each timepoint from

993 independent sample preparations and sequencing, we performed trajectory

994 analysis in pseudo-bulk RNA-seq samples, pooling all cells from each library. For

995 this, valid cell barcodes identified in the previous analysis were collected and

996 their UMI count matrices were retrieved. All UMI count values were then

997 aggregated by genes to generate pseudo-bulk RNA-seq data. We applied Monocle

998 2 (Qiu et al., 2017; Trapnell et al., 2014) for this trajectory analysis using 2758

999 highly variable genes or top 500 most differentially expressed (DE) genes out of

1000 7596 genes (expressing more than one UMI in at least three cells) with default

1001 parameters. The highly variable genes were selected with criteria of

1002 *dispersion_empirical >= 1 * dispersion_fit & mean_expression >= 1*. In both analyses,

1003 the results were similar, so the trajectory using variable genes was presented

1004 (Figure S1E).

1005

Subclustering analysis

1006

1007   To determine detailed cellular states or subtypes and to exclude

1008 unintended cells that originate from only a single library, each of eight major cell

1009 types was clustered separately using Seurat. For this, 1) cells designated to each

1010 cell type were retrieved from the integrated Seurat object, 2) their UMI counts

1011 were scaled and normalized for sequencing library and total UMI counts, and 3)

1012 the number of PCs were determined for dimension reduction and clustering

1013 analysis. In each subclustering analysis, hundreds of t-SNE and UMAP plots were

1014 manually examined with random seeds for visualization. Resolution for

1015 clustering analysis was manually selected as follows: 13 subclusters for PH (*res* =

1016 0.8), 14 for PM (*res* = 0.9), two for LM (*res* = 0.3), three for CC (*res* = 0.1), and

1017 another three for adipohemocytes (*res* = 0.3) were detected. Other cell types

1018 (PSC, GST-rich, and DV) were not further clustered. Next, subclusters mainly

1019 originating from a single library (more than half) were removed from the

1020 datasets because they may have resulted from experimental artefacts. In fact,

1021 two, three, one, and two subclusters were excluded from PH, PM, CC, and

1022 adipohemocyte, correspondingly (Figure S2B).

1023   When the results were examined using known marker genes, one PM

1024 subcluster (PM11) that displayed a high level of ring gland marker gene, *phm*,

1025 was excluded from subsequent analyses. Particularly, of 11 remaining PH

1026 subclusters, one PH subcluster (PH1) displayed a high level of genes related to

1027 the Notch signaling pathway, which is known to be important for prohemocyte

1028 maintenance and differentiation. The PH1 subcluster forms a small group of cells

1029 separated from the main PH cluster in the first clustering analysis, which seems

1030   to be more related to the PH2 subcluster displaying distinct molecular signatures

1031   (Figure 2B). We thus sought to investigate these subclusters in more detail, along

1032   with PSC cells which are known to maintain early prohemocytes in lymph glands.

1033   For this, we reclustered PH1, PH2, and PSC cells together, and found 6

1034   subclusters. However, one of these subclusters appeared to be neuronal cells as

1035   they expressed specific neuronal markers, such as *nSyb* or *Syt1* (Figure S2D). On

1036   the other hand, three PSC subclusters were detected but merged because they all

1037   shared the same signature genes (Figure 3D). In addition, 13 PH cells—

1038   annotated as PSC and 3 PSC cells—that comingled with PH1 or 2 in the

1039   subclustering analysis were removed, as their identities were inconsistent. In

1040   summary, we defined a total of 31 subclusters from the normal lymph gland

1041   dataset (Figure 2); 11 for PH, 10 for PM, two each for LM and CC, one each for

1042   PSC, GST-rich, and adipohemocyte, and three non-hematopoietic cell types (DV,

1043   RG, and Neurons).

1044

1045   **Trajectory analysis using Monocle 3**

1046   To reconstruct lymph gland hematopoiesis in *Drosophila*, hematopoietic

1047   cells in the blood lineage were collected after filtering the PSC, DV, RG, and

1048   neuron subclusters. We then followed the Monocle 3 (Cao et al., 2019) analysis

1049   pipeline described in the website documentation (https://cole-trapnell-

1050   lab.github.io/monocle3/) using custom parameters predetermined from

1051   repetitive analyses. We normalized the dataset by log-transformation and size

1052   factor correction, following three covariates, sequencing library, UMI count, and

1053   mitochondrial gene contents (the proportion of mitochondrial genes in

1054   transcriptome), and scaled using the *preprocess_cds()* function with 75 PCs.

1055   UMAP dimension reduction (*reduce_dimension*) was performed with custom

1056   parameters *umap.min_dist* = 0.4 and *max_components* = 3, and clustering

1057   resolution (*cluster_cells*) was set to 0.001 which assigned all cells into a single

1058   partition. After graph learning was performed (*learn_graph*), the cells were

1059   ordered using *order_cells()* to set a node embedded in the PH1 subcluster as a

1060   start point. All the trajectory graphs were visualized using the *plot_cells()*

1061   function with or without a trajectory graph.

1062   Monocle 3 offers several approaches for differential expression analyses

1063   using regression or graph-autocorrelation. In this study, we identified co-

1064   regulated genes along the pseudotime by graph-autocorrelation and

1065   modularized them. To detect co-regulated genes, the graph-autocorrelation

1066   function *graph_test()* was specified with a "principal_graph" parameter and

1067   significant genes were selected (q value < 0.05). Modularization was performed

1068   using *find_gene_modules()* with default parameters but only passing a list of

1069   resolution values from 10^{-6} to 0.1 for automatic parameter selection. A total of 51

1070   gene modules were detected from the normal lymph gland trajectory, however,

1071   three modules were excluded in that they were unable to be characterized by the

1072   enrichment analysis with biological process gene ontology terms or KEGG

1073   pathways using g:Profiler (https://biit.cs.ut.ee/gprofiler/gost).

1074   The trajectory analysis of infested lymph gland (Figure 5D) followed a

1075   similar pipeline as that of normal lymph gland with slightly different parameters.

1076   The dataset was normalized and corrected for covariates, sequencing library,

1077   UMI count, and mitochondrial gene contents in the same manner, though using

47

1078 50 PCs. We then used 0.5 for the minimum distance in UMAP and 0.005 for the

1079 clustering resolution. Gene modules were also explored using a complete dataset

1080 and the result generally agreed with the previous gene modules using normal

1081 lymph glands. So, we focused on the LM trajectory in the analysis. First, we

1082 filtered all the other subclusters except for PH8, PM1, and PM6 with two LM

1083 subclusters because these subclusters were mainly found in the LM

1084 differentiation paths. Then, we collected cells by excluding those having a UMAP

1085 1 coordination higher than -1.5 (Figure 5E, inset). Modularization of co-regulated

1086 genes was performed as previously described using these cells.

1087

1088 **Investigation of transcription factor activity using SCENIC**

1089 To predict transcription factor regulatory networks in developing lymph

1090 glands, we performed SCENIC (Aibar et al., 2017) analyses on the 1) normal

1091 lymph gland dataset excluding non-hematopoietic cells (n = 19,332) and 2) early

1092 PH subclusters with PSC (PH1, PH2, and PSC; n = 77, 79, and 189, respectively).

1093 We followed the general SCENIC workflow from gene filtration to binarization of

1094 transcription factor (TF) activity described by the authors

1095 (https://www.aertslab.org/#scenic), using a provided cisTarget reference based

1096 on *Drosophila melanogaster* release 6.02. For the latter analysis, we filtered

1097 genes using the following two criteria: 1) genes expressed higher than a UMI

1098 count threshold of 3*(2.5% of total cell count), and 2) genes expressed in at least

1099 2.5% of total cells (8625 cells). We retrieved 4588 genes from these steps and

1100 performed the analysis.

1101   When we investigated the normal lymph gland dataset, however,

1102   predictions for several well known TFs in cell types with relatively small sizes,

1103   such as *lz* for CC (1.44% of the input dataset) or *Antp* for PSC (0.97% of the input

1104   dataset), were affected by the presence of other major cell types. For example, *lz*

1105   was an active TF of CC only when PSC, adipohemoctye and GST-rich cell types

1106   were excluded from the analysis. We reasoned that the relative population size

1107   may affect the predictive power, and signals of TFs active in small populations

1108   would be frequently ignored. To tackle this issue, we randomly sampled 42 cells

1109   (two thirds of the smallest subcluster—LM2, $n = 63$) from each of 28

1110   hematopoietic subclusters, then collected active TFs using a SCENIC workflow

1111   with slightly different gene filtration parameters (we modified 2.5% of the total

1112   cell count to 1.0% of total cell count in both gene filtration criteria). We

1113   iteratively performed 100 independent trials measuring frequencies for TFs and

1114   collected 45 out of 177 TFs predicted active at least 25 times. Then we generated

1115   pseudo-bulk profiles for cell types in each timepoint by aggregating scaled gene

1116   expression values retrieved from the Seurat object. During the pseudo-bulk

1117   generation, LM, CC, GST-rich, and adipohemocyte from 72 h AEL ($n = 2, 4, 14,$ and

1118   1, respectively) and LM and adipohemocyte from 96 h AEL ($n = 14$ and 5,

1119   respectively) were filtered because the number of cells in these groups was less

1120   than 0.1% (~19 cells) of the complete lymph gland dataset. Using pseudo-bulk

1121   profiles, we produced a heatmap describing TF expression in cell types for each

1122   timepoint (Figure S1K) using the R package *pheatmap*.

1123

1124   **Comparative analysis between normal and wasp infested lymph glands**

49

1125   For the integration of infested lymph gland (24 h PI; 96 h AEL) with

1126   normal dataset from 96 h AEL, those datasets were aligned using Seurat 3.0.

1127   functions, *FindTransferAnchors()* and *TransferData()* with default parameters.

1128   Again, UMI counts were normalized, log-transformed, and scaled to properly

1129   integrate datasets, and the number of PCs was determined (46 PCs).

1130   Subclustering labels defined in normal 96 h AEL lymph gland were transferred to

1131   infested lymph gland, and three minor subclusters were excluded because they

1132   were only found in the normal dataset; PH11 (n = 3), PM10 (n = 3), and

1133   adipohemocyte (n = 5). Three non-hematopoietic cell types (35 DV, 18 neurons,

1134   and 12 RG cells) were not considered in subsequent analysis.

1135

1136   **Comparative analysis between lymph gland and circulation**

1137   A UMI count matrix of 120 h AEL circulating hemocytes described in

1138   Tattikota *et al.* was used. Given that the gene annotation version used in Sudhir

1139   *et al.* was different from ours, an FBgn to Annotation ID conversion table was

1140   downloaded from FlyBase (https://flybase.org;

1141   fbgn_annotation_ID_fb_2019_03.tsv) to match gene IDs. After filtration of non-

1142   hematopoietic cell types (DV, RG) from the lymph gland dataset and circulating

1143   hemocytes expressing high mitochondrial genes (higher than 20%), 17,122 cells

1144   from 96 and 120 h AEL lymph glands (n = 9339 and 7783, respectively) were

1145   aligned with 3706 hemocytes from Tattikota *et al.* dataset and additional 96 and

1146   120 h AEL circulation datasets (n = 1235, 977, and 1494, respectively) using 54

1147   PCs. Subclustering labels in lymph glands were transferred to the circulation

1148   datasets using *FindTransferAnchors()* and *TransferData()* with default

50

1149   parameters. Prior to the comparison of aligned cell types, a number of

1150   subclusters in circulation datasets were filtered because they were either 1)

1151   subclusters mainly originated from a single library (more than half) or 2) minor

1152   subclusters (less than 1% of total circulation cells). After filtration, seven

1153   subclusters remained for subsequent analysis (PH1 = 67, PM4 = 355, PM5 =

1154   1224, PM6 = 961, PM7 = 896, CC1 = 99, and CC2 = 104). To compare gene

1155   expression between circulating and lymph gland hemocytes, each cell type or

1156   subcluster was independently analyzed for differentially expressed genes using

1157   Seurat. Then, gene expression values from each dataset were averaged and log-

1158   transformed and top 10 DEGs were marked for visualization (Figure S6).

1159

1160   **Annotation of HCA bone marrow cells**

1161   The preview dataset produced by HCA, Census of Immune Cells

1162   (Rozenblatt-Rosen et al., 2017), was downloaded from the HCA data portal

1163   (https://preview.data.humancellatlas.org). The dataset was produced from bone

1164   marrow samples donated by eight individuals and sequenced using 10X

1165   Genomics Chromium. To capture high-quality cells from the raw data, we filtered

1166   cells that fail to meet the following criteria: 1) those expressing less than 500

1167   genes, 2) those with UMI counts higher than two standard deviations from mean

1168   UMI counts, 3) those expressing mitochondrial genes higher than 10% of total

1169   gene expression. After the preprocessing step, 262,638 cells were passed to the

1170   subsequent analysis. Cells from different donors were aligned using

1171   *FindIntegrationAnchors()* and *IntegrateData()* with default parameters,

1172   respectively. UMI counts were normalized, log-transformed, and scaled to

1173   properly integrate datasets, and 61 PCs were then used for UMAP projection and

1174 clustering. Clustering was performed with resolution of 0.8 to get 34 clusters. All

1175 the 34 clusters were annotated based on expression of known marker genes.

1176 Clusters were then compared using Spearman correlation of gene expression

1177 values, and similar clusters (Spearman's *rho* ≥ 0.95) were aggregated to obtain

1178 the 19 clusters shown in Figure 7A.

1179

1180 **Gene set variance analysis of *Drosophila* hemocyte signature genes**

1181 Orthologous genes between *Drosophila* and human were searched using

1182 DRSC Integrative Ortholog Prediction Tool (DIOPT) version 8.0 (Hu et al., 2011)

1183 with default settings. The gene expression of 262,638 HCA cells was aggregated

1184 by annotated cell types to obtain a pseudo-bulk matrix, and the non-

1185 hematopoietic stromal cell cluster was excluded. Then, 6463 conserved human

1186 genes which met one of the following criteria were converted to those of the fly:

1187 1) pairs of genes conserved in one-to-one manner or 2) pairs of genes scored

1188 with the highest weighted DIOPT score when conserved in a many-to-one or

1189 one-to-many manner.

1190 Signature genes of seven major hemocytes were analyzed by Seurat DEG

1191 analysis using *FindAllMarkers()* with *min.pct = 0.25* and *only.pos = TRUE*

1192 paramters and only statistically significant genes were left (adjusted *P*-value ≤

1193 0.05; false discovery rate). Then, these genes were searched in fly-human gene

1194 pairs to count the number of signature genes conserved in each major cell type.

1195 PH had the least number of conserved signature genes (*n* = 30), and thus, the top

1196 30 conserved signature genes were used for enrichment analysis in other cell

1197 types. Gene set variance analysis (GSVA) (Hanzelmann et al., 2013) was

1198 performed on pseudo-bulk HCA data with these customized gene sets to identify

1199   the enrichment of hemocyte signatures in human bone marrow. GSVA scores

1200   were clustered and visualized using an R package *pheatmap* (Figure 7C).

1201

1202   **Bulk RNA-seq of the lymph gland**

1203   Procedures prior to the Papain treatment were performed to collect intact

1204   lymph glands. Instead of Papain solution, TRIzol (MRC, TR118) was added to

1205   lymph glands and RNA extraction performed. More than 1ug of RNA was

1206   prepared for each experiment. A cDNA library was constructed by 5' and 3'

1207   adapter ligation and loaded into a flow cell where fragments are captured into

1208   library adapters. Each fragment was then amplified through bridge amplification.

1209   After cluster generation was completed, templates were sequenced by Illumina

1210   TruSeq.

1211

1212   **Wasp Infestation**

1213   Larvae were infested at 72 h AEL with *Leptopilina boulardi*. Wasps were

1214   removed after 8 hours of co-culture and egg deposition was confirmed by direct

1215   observation of wasp eggs in the hemolymph during dissection. All infestation

1216   procedures were performed at 25 °C.

1217

1218   **Immunohistochemistry**

1219   Lymph glands were dissected and stained as previously described (Jung

1220   et al., 2005). The following primary antibodies were used in this study: α-lz

1221   (DSHB, 1:10), α-Antp (DSHB, 1:10), α-Dl (DSHB, 1:10), α-L1 (I.Ando, 1:100), α-

1222   col (M.Crozatier, 1:400), α-Ubx (DSHB, 1:10), α-nc82 (DSHB, 1:10), α-NimC1

1223   (I.Ando, 1:100), α-Pxn (Yoon et al., 2017), α-GFP (Sigma Aldrich; G6539; 1:2000)

53

1224   and α-F-actin (ThermoFisher; A34055, 1:200). Cy3-, FITC- or Alexa Fluor 647-

1225   conjugated secondary antibody (Jackson Laboratory) was used for staining at a

1226   1:250 ratio. All samples were kept in VectaShield (Vector Laboratory) and

1227   imaged by a Nikon C2 Si-plus confocal microscope.

1228   For α-Delta staining, a pre-absorption step was essential to reduce

1229   background in the lymph gland. For the pre-absorption step, 1:10 diluted (with

1230   2% sodium azide) α-Dl antibody was incubated together with 9 fixed larval

1231   cuticles overnight at 4°C. Lymph glands were dissected in Schneider's medium

1232   and cultured in Schneider's medium with 10mM EDTA for 30 minutes. Samples

1233   were incubated with 3.7% formaldehyde for 1 hour for fixation at 25 °C. After

1234   fixation, lymph glands were washed 3 times in 0.1% Triton X-100 in 1xPBS and

1235   blocked in 1% BSA/0.1% Triton X in 1xPBS for 1 hour. Lymph glands were

1236   incubated overnight at 4°C with α-Dl antibody. Lymph glands were washed 3

1237   times in 0.1% Triton X in 1xPBS and then incubated with α-mouse secondary

1238   antibody with 1% BSA/0.1% Triton X in 1xPBS for 3 hours at room temperature.

1239   After washing 3 times with 0.1% Triton X in 1xPBS, samples were mounted in

1240   Vectashield (Vector Laboratory) with DAPI and imaged by a Nikon C2 Si-plus

1241   confocal microscope.

1242

1243   **Bleeding hemocytes**

1244   Larvae were vortexed with glass beads (Sigma G9268) for one minute

1245   before bleeding to detach sessile hemocytes (Petraki et al., 2015). Larvae were

1246   bled on a slide glass (Immuno-Cell Int.; 61.100.17) and hemocytes allowed to

1247   settle onto the slide at 4 °C for 40 minutes. Hemocytes were washed 3 times in

54

1248   0.4% Triton X-100 in 1x PBS for 10 minutes and blocked in 1% BSA/0.4%

1249   TritonX in 1xPBS for 30 minutes. Primary antibody was added and samples

1250   incubated overnight at 4°C. Hemocytes were washed 3 times in 0.4% Triton X in

1251   1xPBS and then incubated with a secondary antibody with 1% BSA/0.4% Triton

1252   X in 1xPBS for 3 hours at room temperature. After washing 3 times with 0.4%

1253   Triton X in 1xPBS, samples were kept in Vectashield (Vector Laboratory) with

1254   DAPI and imaged by a Nikon C2 Si-plus confocal microscope.

1255

1256   **Quantification of samples**

1257   Stained or fluorescent cells were quantified and analyzed by IMARIS

1258   software (Bitplane). Individual primary lobes were counted for this study. *In vivo*

1259   data was analyzed by the Wilcoxon rank sum test after determining normality

1260   with the use of SPSS (version 24).

1261

1262   **Fluorescent *in situ* hybridization**

1263   An *in situ* hybridization protocol in a previous study (Crozatier et al.,

1264   1996) was used. An anti-*delta* probe was designed based on *delta* cDNA

1265   sequences (Forward primer: ATGTGCGAGGAGAAAGTGCT, Reverse primer:

1266   CGACTTGTCCCAGGTGTTTT). DIG-labeled probes were detected by α-DIG-biotin

1267   antibody (Jackson Immunoresearch; 200-062-156) and visualization was done

1268   using a SuperBoost™ Kit (ThermoFisher; B40933). The sense probe was used as

1269   a negative control.

1270

1271   **Data and software availability**

55

1272   In-house R and Python codes that were implemented in this study are

1273   available on GitHub (https://github.com/sangho1130/Dmel_Dropseq). Raw

1274   scRNA-seq and bulk RNA-seq reads are available through the NCBI Gene

1275   Expression Omnibus (GEO) (GSE141275).

Reference

Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., *et al.* (2017). SCENIC: single-cell regulatory network inference and clustering. Nat Methods *14*, 1083-1086.

Anderl, I., Vesala, L., Ihalainen, T.O., Vanha-Aho, L.M., Ando, I., Ramet, M., and Hultmark, D. (2016). Transdifferentiation and Proliferation in Two Distinct Hemocyte Lineages in Drosophila melanogaster Larvae after Wasp Infection. PLoS Pathog *12*, e1005746.

Baldeosingh, R., Gao, H., Wu, X., and Fossett, N. (2018). Hedgehog signaling from the Posterior Signaling Center maintains U-shaped expression and a prohemocyte population in Drosophila. Dev Biol *441*, 132-145.

Banerjee, U., Girard, J.R., Goins, L.M., and Spratford, C.M. (2019). Drosophila as a Genetic Model for Hematopoiesis. Genetics *211*, 367-417.

Bazzi, W., Cattenoz, P.B., Delaporte, C., Dasari, V., Sakr, R., Yuasa, Y., and Giangrande, A. (2018). Embryonic hematopoiesis modulates the inflammatory response and larval hematopoiesis in Drosophila. Elife *7*.

Benmimoun, B., Polesello, C., Haenlin, M., and Waltzer, L. (2015). The EBF transcription factor Collier directly promotes Drosophila blood cell progenitor maintenance independently of the niche. Proc Natl Acad Sci U S A *112*, 9052-9057.

Binggeli, O., Neyen, C., Poidevin, M., and Lemaitre, B. (2014). Prophenoloxidase activation is required for survival to microbial infections in Drosophila. PLoS Pathog *10*, e1004067.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment *2008*, P10008.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol *36*, 411-420.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., *et al.* (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496-502.

Cooper, E.L. (1976). Evolution of blood cells. Ann Immunol (Paris) *127*, 817-825.

Crozatier, M., Ubeda, J.M., Vincent, A., and Meister, M. (2004). Cellular immune response to parasitization in Drosophila requires the EBF orthologue collier. PLoS Biol *2*, E196.

Crozatier, M., Valle, D., Dubois, L., Ibnsouda, S., and Vincent, A. (1996). Collier, a novel regulator of Drosophila head development, is expressed in a single mitotic domain. Curr Biol *6*, 707-718.

Crozatier, M., and Vincent, A. (2011). Drosophila: a model for studying genetic and molecular aspects of haematopoiesis and associated leukaemias. Dis Model Mech *4*, 439-445.

Dey, N.S., Ramesh, P., Chugh, M., Mandal, S., and Mandal, L. (2016). Dpp dependent Hematopoietic stem cells give rise to Hh dependent blood progenitors in larval lymph gland of Drosophila. Elife *5*.

1322 Edgar, B.A., Lehman, D.A., and O'Farrell, P.H. (1994). Transcriptional regulation
1323 of string (cdc25): a link between developmental programming and the cell cycle.
1324 Development *120*, 3131-3143.

1325 Evans, C.J., Hartenstein, V., and Banerjee, U. (2003). Thicker than blood:
1326 conserved mechanisms in Drosophila and vertebrate hematopoiesis. Dev Cell *5*,
1327 673-690.

1328 Evans, C.J., Liu, T., and Banerjee, U. (2014). Drosophila hematopoiesis: Markers
1329 and methods for molecular genetic analysis. Methods *68*, 242-251.

1330 Ferguson, G.B., and Martinez-Agosto, J.A. (2014). Yorkie and Scalloped signaling
1331 regulates Notch-dependent lineage specification during Drosophila
1332 hematopoiesis. Curr Biol *24*, 2665-2672.

1333 Ferguson, G.B., and Martinez-Agosto, J.A. (2017). The TEAD family transcription
1334 factor Scalloped regulates blood progenitor maintenance and proliferation in
1335 Drosophila through PDGF/VEGFR receptor (Pvr) signaling. Dev Biol *425*, 21-32.

1336 Franc, N.C., Dimarcq, J.L., Lagueux, M., Hoffmann, J., and Ezekowitz, R.A. (1996).
1337 Croquemort, a novel Drosophila hemocyte/macrophage receptor that recognizes
1338 apoptotic cells. Immunity *4*, 431-443.

1339 Gajewski, K.M., Sorrentino, R.P., Lee, J.H., Zhang, Q., Russell, M., and Schulz, R.A.
1340 (2007). Identification of a crystal cell-specific enhancer of the black cells
1341 prophenoloxidase gene in Drosophila. Genesis *45*, 200-207.

1342 Galloway, J.L., and Zon, L.I. (2003). Ontogeny of hematopoiesis: examining the
1343 emergence of hematopoietic cells in the vertebrate embryo. Curr Top Dev Biol
1344 *53*, 139-158.

1345 Ghosh, S., Singh, A., Mandal, S., and Mandal, L. (2015). Active hematopoietic hubs
1346 in Drosophila adults generate hemocytes and contribute to immune response.
1347 Dev Cell *33*, 478-488.

1348 Gold, K.S., and Bruckner, K. (2014). Drosophila as a model for the two myeloid
1349 blood cell systems in vertebrates. Exp Hematol *42*, 717-727.

1350 Grigorian, M., Mandal, L., and Hartenstein, V. (2011). Hematopoiesis at the onset
1351 of metamorphosis: terminal differentiation and dissociation of the Drosophila
1352 lymph gland. Dev Genes Evol *221*, 121-131.

1353 Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation
1354 analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

1355 Hao, Y., and Jin, L.H. (2017). Dual role for Jumu in the control of hematopoietic
1356 progenitors in the Drosophila lymph gland. Elife *6*.

1357 Hartenstein, V. (2006). Blood cells and blood cell development in the animal
1358 kingdom. Annu Rev Cell Dev Biol *22*, 677-712.

1359 Hillyer, J.F., Schmidt, S.L., and Christensen, B.M. (2003). Hemocyte-mediated
1360 phagocytosis and melanization in the mosquito Armigeres subalbatus following
1361 immune challenge by bacteria. Cell Tissue Res *313*, 117-127.

1362 Holz, A., Bossinger, B., Strasser, T., Janning, W., and Klapper, R. (2003). The two
1363 origins of hemocytes in Drosophila. Development *130*, 4955-4962.

1364 Honti, V., Csordas, G., Kurucz, E., Markus, R., and Ando, I. (2014). The cell-
1365 mediated immunity of Drosophila melanogaster: hemocyte lineages, immune
1366 compartments, microanatomy and regulation. Dev Comp Immunol *42*, 47-56.

1367 Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and
1368 Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-
1369 focused and other functional studies. BMC Bioinformatics *12*, 357.

Irving, P., Ubeda, J.M., Doucet, D., Troxler, L., Lagueux, M., Zachary, D., Hoffmann, J.A., Hetru, C., and Meister, M. (2005). New insights into Drosophila larval haemocyte functions through genome-wide analysis. Cell Microbiol *7*, 335-350.

Jung, S.H., Evans, C.J., Uemura, C., and Banerjee, U. (2005). The Drosophila lymph gland as a developmental model of hematopoiesis. Development *132*, 2521-2533.

Kocks, C., Cho, J.H., Nehme, N., Ulvila, J., Pearson, A.M., Meister, M., Strom, C., Conto, S.L., Hetru, C., Stuart, L.M.*, et al.* (2005). Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in Drosophila. Cell *123*, 335-346.

Kroeger, P.T., Jr., Tokusumi, T., and Schulz, R.A. (2012). Transcriptional regulation of eater gene expression in Drosophila blood cells. Genesis *50*, 41-49.

Krzemien, J., Dubois, L., Makki, R., Meister, M., Vincent, A., and Crozatier, M. (2007). Control of blood cell homeostasis in Drosophila larvae by the posterior signalling centre. Nature *446*, 325-328.

Krzemien, J., Oyallon, J., Crozatier, M., and Vincent, A. (2010). Hematopoietic progenitors and hemocyte lineages in the Drosophila lymph gland. Dev Biol *346*, 310-319.

Kurucz, E., Markus, R., Zsamboki, J., Folkl-Medzihradszky, K., Darula, Z., Vilmos, P., Udvardy, A., Krausz, I., Lukacsovich, T., Gateff, E.*, et al.* (2007). Nimrod, a putative phagocytosis receptor with EGF repeats in Drosophila plasmatocytes. Curr Biol *17*, 649-654.

Lanot, R., Zachary, D., Holder, F., and Meister, M. (2001). Postembryonic hematopoiesis in Drosophila. Dev Biol *230*, 243-257.

Lebestky, T., Chang, T., Hartenstein, V., and Banerjee, U. (2000). Specification of Drosophila hematopoietic lineage by conserved transcription factors. Science *288*, 146-149.

Lebestky, T., Jung, S.H., and Banerjee, U. (2003). A Serrate-expressing signaling center controls Drosophila hematopoiesis. Genes Dev *17*, 348-353.

Leitao, A.B., and Sucena, E. (2015). Drosophila sessile hemocyte clusters are true hematopoietic tissues that regulate larval blood cell differentiation. Elife *4*.

Lemaitre, B., and Hoffmann, J. (2007). The host defense of Drosophila melanogaster. Annu Rev Immunol *25*, 697-743.

Li, A.C., and Glass, C.K. (2002). The macrophage foam cell as a target for therapeutic intervention. Nat Med *8*, 1235-1242.

Llamazares, S., Moreira, A., Tavares, A., Girdham, C., Spruce, B.A., Gonzalez, C., Karess, R.E., Glover, D.M., and Sunkel, C.E. (1991). polo encodes a protein kinase homolog required for mitosis in Drosophila. Genes Dev *5*, 2153-2165.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M.*, et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Madhwal, S., Shin, M., Joshi, M.K., Kapoor, A., Ur Rehman, P.M., Gor, K., Shim, J., and Mukherjee, T. (2019). Metabolic control of immune-competency by odors in Drosophila. bioRxiv, 718056.

Makhijani, K., Alexander, B., Tanaka, T., Rulifson, E., and Bruckner, K. (2011). The peripheral nervous system supports blood cell homing and survival in the Drosophila larva. Development *138*, 5379-5391.

Mandal, L., Banerjee, U., and Hartenstein, V. (2004). Evidence for a fruit fly hemangioblast and similarities between lymph-gland hematopoiesis in fruit fly and mammal aorta-gonadal-mesonephros mesoderm. Nat Genet *36*, 1019-1023.

1419 Mandal, L., Martinez-Agosto, J.A., Evans, C.J., Hartenstein, V., and Banerjee, U.
1420 (2007). A Hedgehog- and Antennapedia-dependent niche maintains Drosophila
1421 haematopoietic precursors. Nature *446*, 320-324.

1422 Markus, R., Laurinyecz, B., Kurucz, E., Honti, V., Bajusz, I., Sipos, B., Somogyi, K.,
1423 Kronhamn, J., Hultmark, D., and Ando, I. (2009). Sessile hemocytes as a
1424 hematopoietic compartment in Drosophila melanogaster. Proc Natl Acad Sci U S
1425 A *106*, 4805-4809.

1426 Mathieu, J., Cauvin, C., Moch, C., Radford, S.J., Sampaio, P., Perdigoto, C.N.,
1427 Schweisguth, F., Bardin, A.J., Sunkel, C.E., McKim, K.*, et al.* (2013). Aurora B and
1428 cyclin B have opposite effects on the timing of cytokinesis abscission in
1429 Drosophila germ cells and in vertebrate somatic cells. Dev Cell *26*, 250-265.

1430 Minakhina, S., and Steward, R. (2010a). Hematopoietic stem cells in
1431 Drosophila. Development *137*, 27-31.

1432 Minakhina, S., and Steward, R. (2010b). Hematopoietic stem cells in Drosophila.
1433 Development *137*, 27-31.

1434 Mondal, B.C., Mukherjee, T., Mandal, L., Evans, C.J., Sinenko, S.A., Martinez-Agosto,
1435 J.A., and Banerjee, U. (2011). Interaction between differentiating cell- and niche-
1436 derived signals in hematopoietic progenitor maintenance. Cell *147*, 1589-1600.

1437 Moore, K.J., Sheedy, F.J., and Fisher, E.A. (2013). Macrophages in atherosclerosis:
1438 a dynamic balance. Nature Reviews Immunology *13*, 709-721.

1439 Nagarkar-Jaiswal, S., Lee, P.T., Campbell, M.E., Chen, K., Anguiano-Zarate, S.,
1440 Gutierrez, M.C., Busby, T., Lin, W.W., He, Y., Schulze, K.L.*, et al.* (2015). A library of
1441 MiMICs allows tagging of genes and reversible, spatial and temporal knockdown
1442 of proteins in Drosophila. Elife *4*.

1443 Owusu-Ansah, E., and Banerjee, U. (2009). Reactive oxygen species prime
1444 Drosophila haematopoietic progenitors for differentiation. Nature *461*, 537-541.

1445 Parry, D.H., and O'Farrell, P.H. (2001). The schedule of destruction of three
1446 mitotic cyclins can dictate the timing of events during exit from mitosis. Curr Biol
1447 *11*, 671-683.

1448 Petraki, S., Alexander, B., and Bruckner, K. (2015). Assaying Blood Cell
1449 Populations of the Drosophila melanogaster Larva. J Vis Exp.

1450 Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017).
1451 Reversed graph embedding resolves complex single-cell trajectories. Nat
1452 Methods *14*, 979-982.

1453 Regan, J.C., Brandao, A.S., Leitao, A.B., Mantas Dias, A.R., Sucena, E., Jacinto, A., and
1454 Zaidman-Remy, A. (2013). Steroid hormone signaling is essential to regulate
1455 innate immune cells and fight bacterial infection in Drosophila. PLoS Pathog *9*,
1456 e1003720.

1457 Rizki, R.M., and Rizki, T.M. (1984). Selective destruction of a host blood cell type
1458 by a parasitoid wasp. Proc Natl Acad Sci U S A *81*, 6154-6158.

1459 Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A., and Teichmann, S.A. (2017).
1460 The Human Cell Atlas: from vision to reality. Nature *550*, 451-453.

1461 Sadasivam, S., and DeCaprio, J.A. (2013). The DREAM complex: master
1462 coordinator of cell cycle-dependent gene expression. Nat Rev Cancer *13*, 585-
1463 595.

1464 Sakaguchi, S., Miyara, M., Costantino, C.M., and Hafler, D.A. (2010). FOXP3+
1465 regulatory T cells in the human immune system. Nat Rev Immunol *10*, 490-500.

1466 Sanchez Bosch, P., Makhijani, K., Herboso, L., Gold, K.S., Baginsky, R., Woodcock,
1467 K.J., Alexander, B., Kukar, K., Corcoran, S., Jacobs, T.*, et al.* (2019). Adult

1468 Drosophila Lack Hematopoiesis but Rely on a Blood Cell Reservoir at the
1469 Respiratory Epithelia to Relay Infection Signals to Surrounding Tissues. Dev Cell.

1470 Sharma, S.K., Ghosh, S., Geetha, A.R., Mandal, S., and Mandal, L. (2019). Cell
1471 Adhesion-Mediated Actomyosin Assembly Regulates the Activity of Cubitus
1472 Interruptus for Hematopoietic Progenitor Maintenance in Drosophila. Genetics
1473 *212*, 1279-1300.

1474 Shim, J., Mukherjee, T., Mondal, B.C., Liu, T., Young, G.C., Wijewarnasuriya, D.P.,
1475 and Banerjee, U. (2013). Olfactory control of blood progenitor maintenance. Cell
1476 *155*, 1141-1153.

1477 Sorrentino, R.P., Carton, Y., and Govind, S. (2002). Cellular Immune Response to
1478 Parasite Infection in the Drosophila Lymph Gland Is Developmentally Regulated.
1479 Developmental Biology *243*, 65-80.

1480 Sorrentino, R.P., Tokusumi, T., and Schulz, R.A. (2007). The Friend of GATA
1481 protein U-shaped functions as a hematopoietic tumor suppressor in Drosophila.
1482 Dev Biol *311*, 311-323.

1483 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd,
1484 Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive
1485 Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

1486 Tang, H., Kambris, Z., Lemaitre, B., and Hashimoto, C. (2006). Two proteases
1487 defining a melanization cascade in the immune system of Drosophila. J Biol Chem
1488 *281*, 28097-28104.

1489 Tepass, U., Fessler, L.I., Aziz, A., and Hartenstein, V. (1994). Embryonic origin of
1490 hemocytes and their relationship to cell death in Drosophila. Development *120*,
1491 1829-1837.

1492 Tikhonova, A.N., Dolgalev, I., Hu, H., Sivaraj, K.K., Hoxha, E., Cuesta-Dominguez, A.,
1493 Pinho, S., Akhmetzyanova, I., Gao, J., Witkowski, M*., et al.* (2019). The bone
1494 marrow microenvironment at single-cell resolution. Nature *569*, 222-228.

1495 Tokusumi, T., Sorrentino, R.P., Russell, M., Ferrarese, R., Govind, S., and Schulz,
1496 R.A. (2009). Characterization of a lamellocyte transcriptional enhancer located
1497 within the misshapen gene of Drosophila melanogaster. PLoS One *4*, e6429.

1498 Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J.,
1499 Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of
1500 cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat
1501 Biotechnol *32*, 381-386.

1502 van der Maaten, L.J.P., and Hinton, G.E. (2008). Visualizing High-Dimensional
1503 Data Using t-SNE. Journal of Machine Learning Research *9*, 2579 - 2605.

1504 Weissman, I.L., Anderson, D.J., and Gage, F. (2001). Stem and progenitor cells:
1505 origins, phenotypes, lineage commitments, and transdifferentiations. Annu Rev
1506 Cell Dev Biol *17*, 387-403.

1507 Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational
1508 Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst *8*,
1509 281-291 e289.

1510 Xavier, M.J., and Williams, M.J. (2011). The Rho-family GTPase Rac1 regulates
1511 integrin localization in Drosophila immunosurveillance cells. PLoS One *6*,
1512 e19504.

1513 Yang, H., Kronhamn, J., Ekstrom, J.O., Korkut, G.G., and Hultmark, D. (2015).
1514 JAK/STAT signaling in Drosophila muscles controls the cellular immune
1515 response against parasitoid infection. EMBO Rep *16*, 1664-1672.

1516   Yoon, S., Cho, B., Shin, M., Koranteng, F., Cha, N., and Shim, J. (2017). Iron
1517   Homeostasis Controls Myeloid Blood Cell Differentiation in Drosophila. Mol Cells
1518   *40*, 976-985.
1519   Yu, S., Luo, F., and Jin, L.H. (2018). The Drosophila lymph gland is an ideal model
1520   for studying hematopoiesis. Dev Comp Immunol *83*, 60-69.
1521   Zielke, N., and Edgar, B.A. (2015). FUCCI sensors: powerful new tools for analysis
1522   of cell proliferation. WIREs Developmental Biology *4*, 469-487.
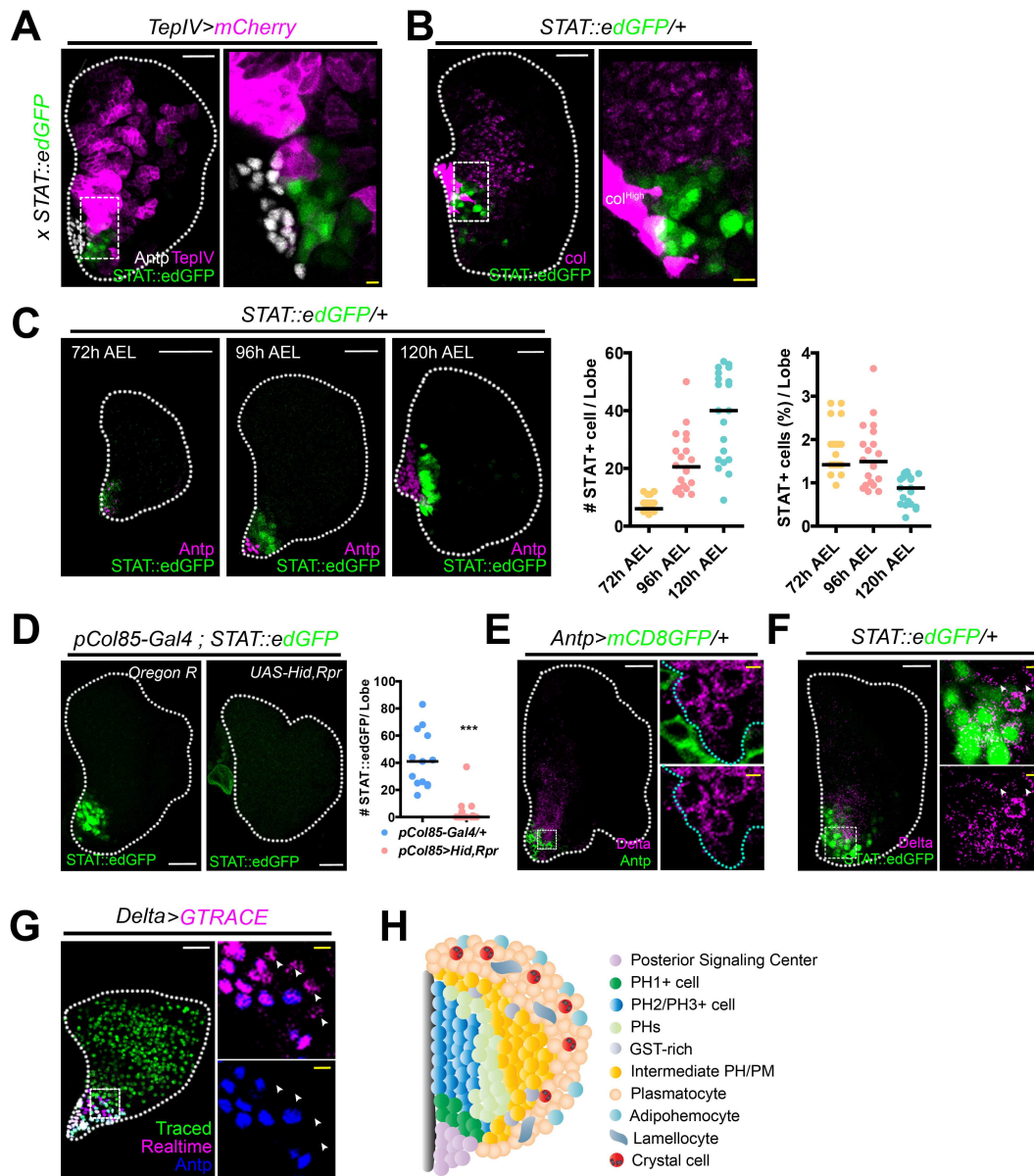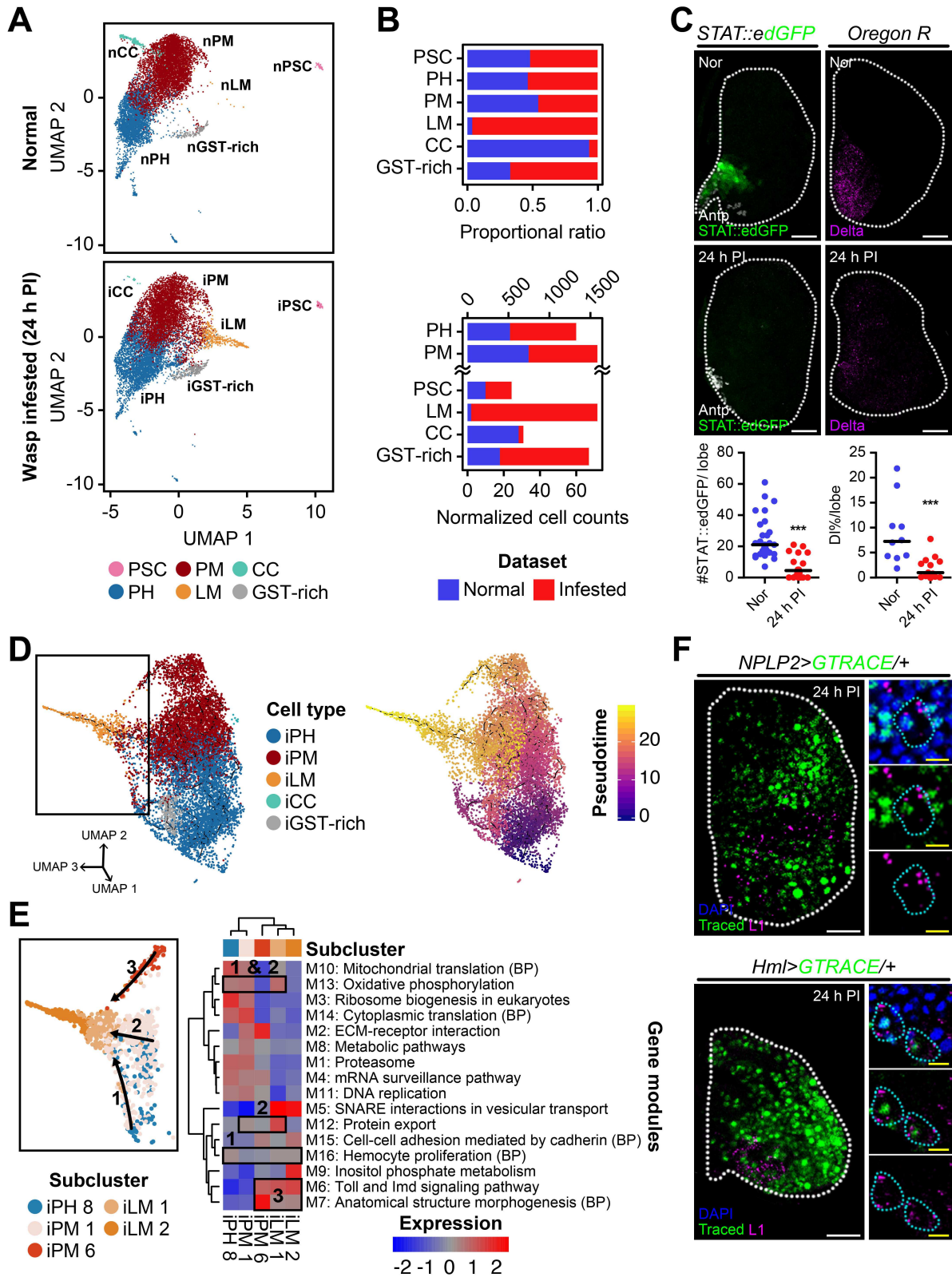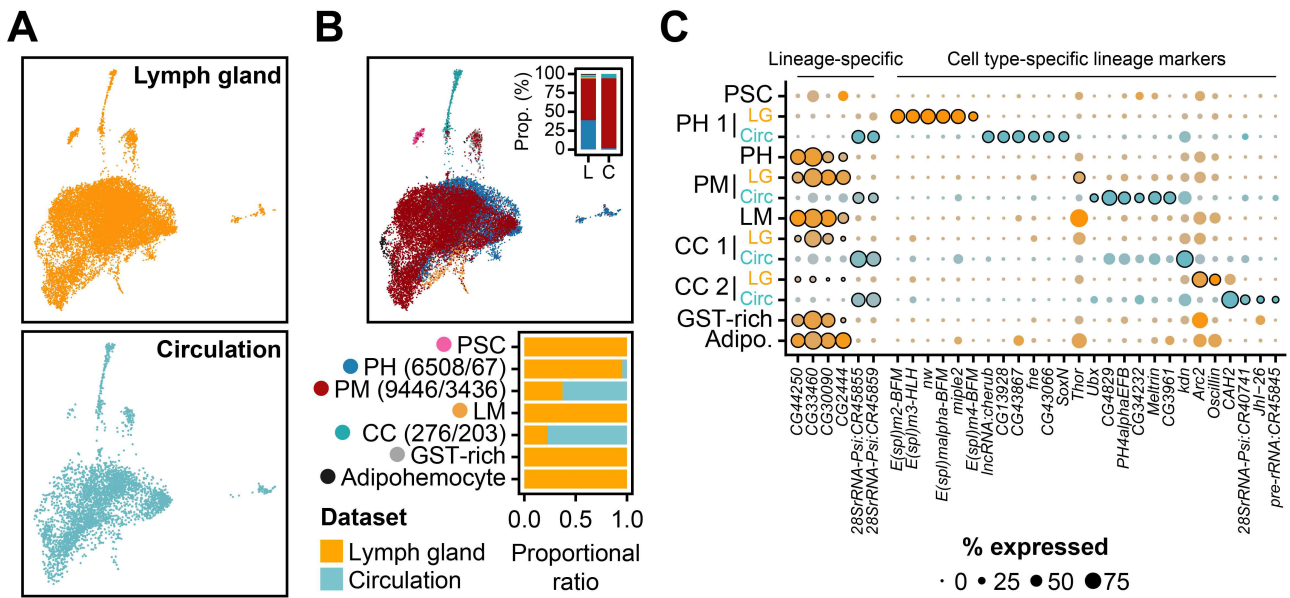1523

Figure 1
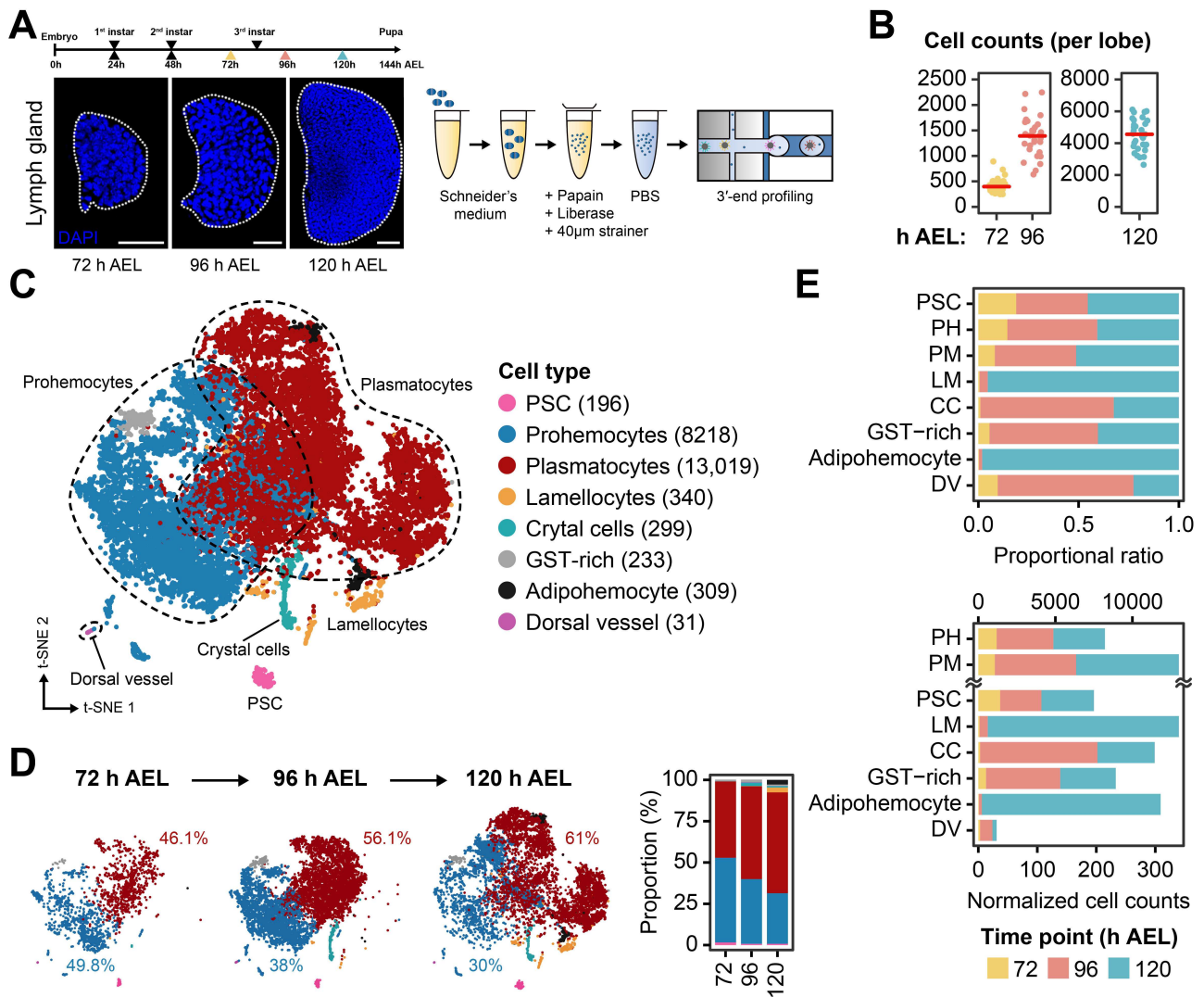
Figure 2
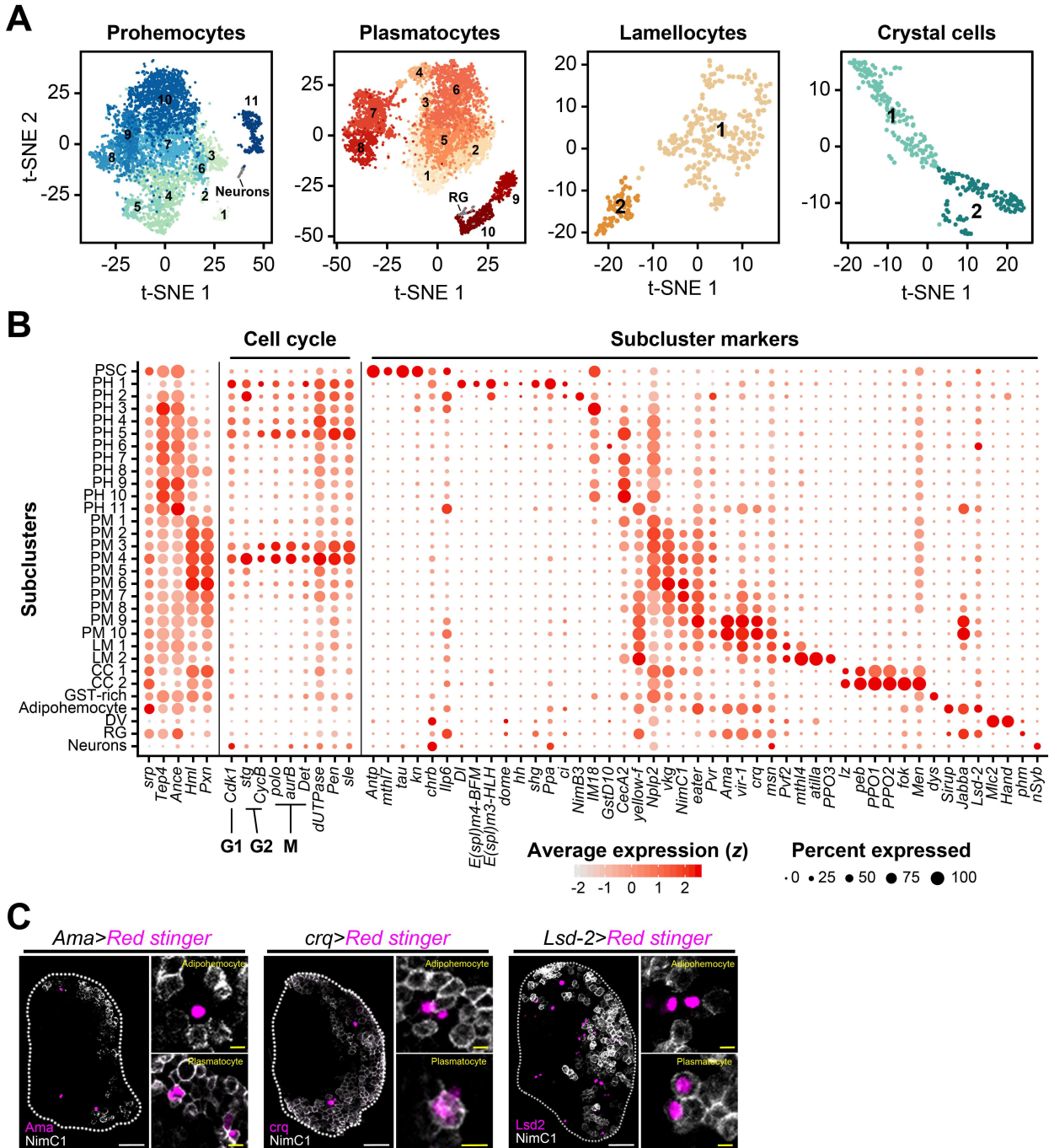
Figure 3
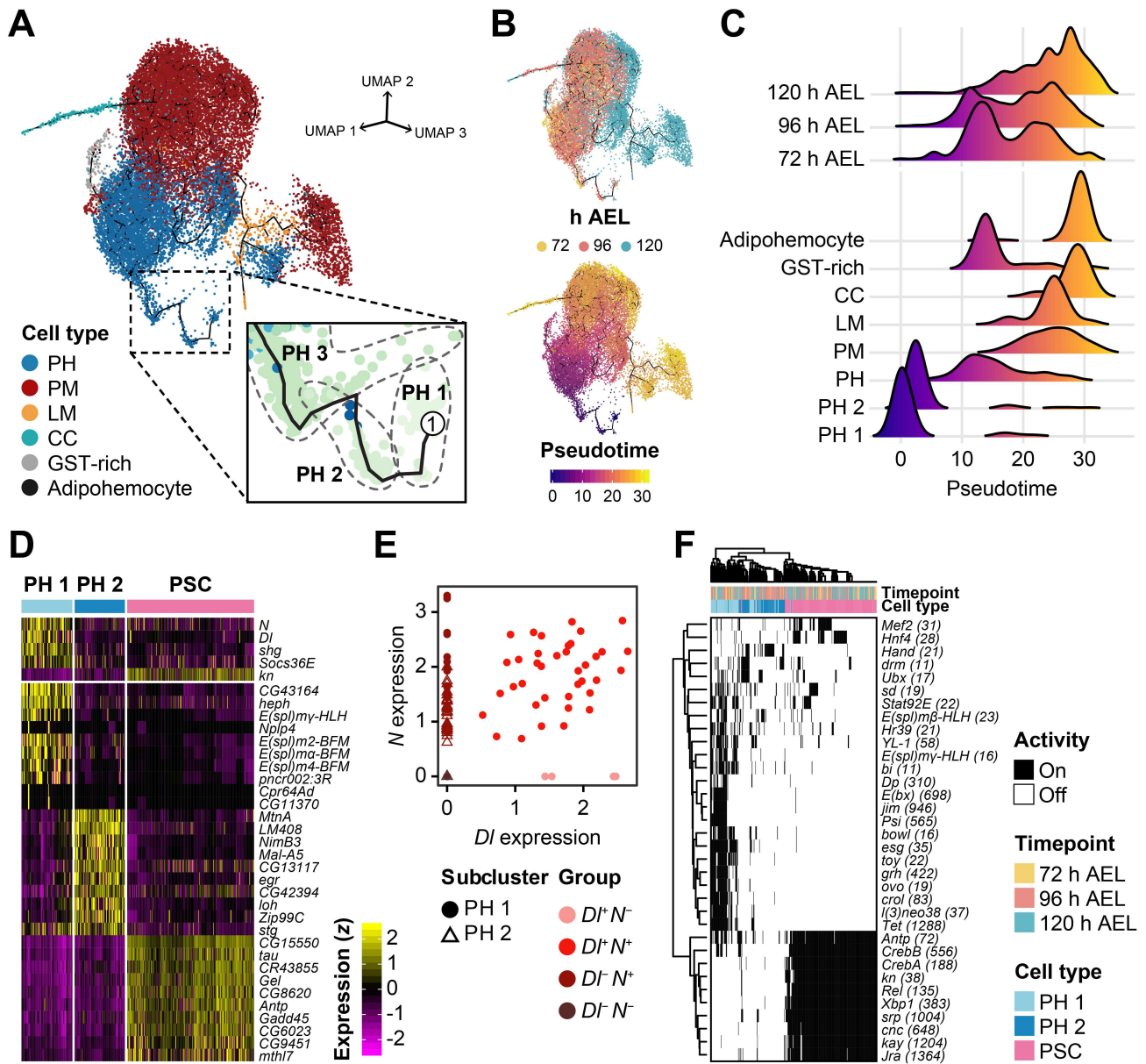
Figure 4

Figure 5

Figure 6

Figure 7