A single cell atlas reveals unanticipated cell type complexity in *Drosophila* ovaries

Maija Slaidina[1], Selena Gupta[1], Torsten U. Banisch[1], Ruth Lehmann[1,2]

1: HHMI, Skirball Institute of Biomolecular Medicine, Department of Cell Biology, NYU Grossman School of Medicine, NY, USA, 540 First Avenue, NYC, NY 10016

2: Present: Whitehead Institute and Department of Biology at MIT, 455 Mainstreet, Cambridge, MA 02142

Abstract

Organ function relies on the spatial organization and functional coordination of numerous cell types. The *Drosophila* ovary is a widely used model system to study the cellular activities underlying organ function, including stem cell regulation, cell signaling and epithelial morphogenesis. However, the relative paucity of cell type specific reagents hinders investigation of molecular functions at the appropriate cellular resolution.

Here, we used single cell RNA sequencing to characterize all cell types of the stem cell compartment and early follicles of the *Drosophila* ovary. We computed transcriptional signatures and identified specific markers for nine states of germ cell differentiation, and 23 somatic cell types and subtypes. We uncovered an unanticipated diversity of escort cells, the somatic cells that directly interact with differentiating germline cysts. Three escort cell subtypes reside in discrete anatomical positions, and express distinct sets of secreted and transmembrane proteins, suggesting that diverse micro-environments support the progressive differentiation of germ cells. Finally, we identified 17 follicle cell subtypes, and characterized their transcriptional profiles. Altogether, we provide a comprehensive resource of gene expression, cell type specific markers, spatial coordinates and functional predictions for 34 ovarian cell types and subtypes.

Introduction

Most organs are composed of numerous cell types. Their spatial organization determines the cellular interactions that ensure sustainable organ function, which lasts for the duration of an organism's life. The fly ovary is an attractive and widely used model system to study the general principles of organ function, due to a combination of genetic tractability, small size, and cell type complexity. Oogenesis, or egg production, is the major function of the ovary. Adult ovaries comprise cells of two distinct lineages: the somatic cells of mesodermal origin, and the germ cells (GCs) that arise from primordial germ cells and differentiate into eggs (Boyle and DiNardo,

37 1995; Moore et al., 1998, Riechmann et al., 1998). Ovaries are composed of 16 to 20 units
38 called ovarioles, each of which serving as an egg production line (Fig. 1C) (Spradling 1993). At
39 the anterior tip of an ovariole is a stem cell compartment, called the germarium, which houses
40 germline stem cells (GSC), and at a more posterior position the somatic follicle stem cells (FSCs)
41 and their support cells (Duhart et al., 2017; Fuller and Spradling, 2007; Kirilly and Xie, 2007).
42 More posteriorly, follicles, or egg chambers, of increasingly advanced stages are lined up, with
43 the mature oocytes located distally. Each follicle contains a germline cyst surrounded by a
44 follicular epithelium, and will produce an egg. Specialized follicle cells, called stalk cells,
45 separate follicles from one another, while polar cells at each pole of the follicle serve as
46 organizers (King and King, 1970; Margaritis et al., 1980, Duhart et al., 2017, Mahowald 1980).

47 Both stem cell populations are excellent models to study stem cell maintenance and
48 differentiation, and stem cell-niche interactions. GSC self-renewal has been extensively studied,
49 and relies on signals from and adhesion to the niche, which comprise terminal filament (TF) and
50 cap cells (CC) (Fig. 3B) (Sahut-Barnola et al., 1995; Song et al., 2002; Xie and Spradling, 1998;
51 2000). On the other hand, the mechanisms governing FSC self-renewal and differentiation are
52 poorly understood (Margolis and Spradling, 1995; Reilein et al., 2017; Rust and Nystul, 2020).

53 Numerous somatic cell types support the multi-stage process of GC differentiation into mature
54 eggs. Specifically, escort cells (ECs), also called inner germarium sheath cells (IGS), send
55 protrusions around GCs and have a dual role: to promote both GSC renewal and GC
56 differentiation (Kirilly et al., 2011; Rojas-Rios et al., 2012; Liu et al., 2015; Wang and Page-
57 McCaw, 2018). This regulation relies on large surface contacts between the two cell types, and
58 it is unclear whether distinct EC populations execute each function (Banisch et al., 2017; Kirilly
59 et al., 2011; Schulz et al., 2002). Oogenesis is further supported by follicle cells (FCs) that arise
60 from FSCs (Margolis and Spradling, 1995). FCs form an epithelial layer around the germline-
61 derived oocyte and nurse cells, thus delimiting the follicles. FCs are essential for egg
62 development; they pre-pattern the body axes of the future embryo, and secrete yolk proteins
63 and the eggshell. Numerous signaling pathways are involved in specification and
64 morphogenesis of follicle cell subtypes (Duhart et al., 2017; Wu et al., 2008, Rust and Nystul,

65 2020). Therefore, the fly ovary also serves as a powerful experimental paradigm to dissect the
66 molecular basis of cell-cell signaling. Despite the identification of various genes regulating
67 oogenesis, the persistent dearth of cell type specific markers and tools has hindered progress in
68 determining the exact role of these factors with the appropriate cellular resolution. For
69 example, specific markers for GSCs, FSC and other somatic cell types remain to be identified. To
70 bridge this gap, we set to use a single-cell RNA sequencing (scRNA-seq) approach to
71 systematically identify and characterize all cell types that compose the stem cell compartment
72 and early-to-mid stage egg chambers.

73 Results:

74 scRNA-seq using two distinct methods produce equivalent results.

75 We aimed to obtain and characterize single cell transcriptomes from the germarium and early
76 follicles to identify cell type specific markers, and thus enable further functional studies. We
77 used a nanoliter droplet based 10x Genomics Chromium system for single cell RNA sequencing
78 (Fig. 1A). We performed the experiment in five replicates and selected high quality
79 transcriptomes (Supplemental Text). Despite using two different methods, all five sequencing
80 experiments produced similar results (Supplemental Text, Fig. 1A, Supplemental Fig. S1A,C,D).
81 Data from all replicates were aligned and merged using the Seurat 3 algorithm for library
82 integration (Supplemental Fig. S1A).

83 Our initial coarse clustering revealed six clusters, visualized on a UMAP (Uniform Manifold
84 Approximation and Projection) plot, where each dot represents a single cell transcriptome (Fig.
85 1B). Two clusters expressed the GC marker *vas* (Lasko and Ashburner 1988), and were thus
86 annotated as germ cells (clusters GC I and II) (Fig. 1B,E). The three largest clusters expressed the
87 follicle cell markers *Fas3* and *eya* (Nystul and Spradling 2007, Bai and Montell 2002), and
88 therefore, corresponded to epithelial follicle cells (clusters FC I, II, and III) (Fig. 1B,E). Finally, the
89 last cluster contained somatic cell types of the germarium - TF and CC that express *en*, ECs that
90 express *Wnt4* and *tj* (Forbes et al. 1996, Mottier-Pavie et al. 2016, Li et al. 2003, Kawashima et

al. 2003), early follicle cell lineages that express *Fas3*, *tj* and *eya* (Nystul and Spradling 2007, Li et al. 2003, Kawashima et al. 2003, Bai and Montell 2002), and stalk and polar cell lineages that express *cas* (Chang et al. 2013) and *upd1* (Silver and Montell 2001) (cluster germarium soma) (Fig. 1B,E). We did not detect a separate cluster containing the muscle cells that ensheath the ovarioles (epithelial sheath cells) or peritoneal sheath, suggesting that those cells were lost during dissections or dissociations.

To compare the transcriptional signatures of these clusters, we computed markers for each cluster, and visualized their expression in a heatmap (Fig. 1D, Supplemental Table 1). GC I and GC II clusters shared a fraction of their marker genes, but there was minimal overlap between marker genes of other clusters, indicating, that these cell types are transcriptionally distinct despite shared mesodermal origins. Thus, we produced a high-quality single cell RNA-sequencing dataset of 15,227 cell transcriptomes, which we used to generate ovarian cell type gene expression profiles, transcriptional signatures, and function predictions.

Identification of 9 distinct steps of GC differentiation

Next, we sought to characterize the transcriptional dynamics underlying GC differentiation. GSCs are in close contact with the niche and are marked by a round GC specific organelle called spectrosome (Lin and Spradling 1995). GSCs divide asymmetrically to give rise to a cystoblast (CB), which subsequently undergoes 4 rounds of synchronous divisions with incomplete cytokinesis to produce a 16-cell cyst. Within the cyst, the spectrosome evolves into a fusome that spans all 16 cells (Huynh 2006). The germarium is divided in four regions organized along the anterior-posterior axis - 1, 2a, 2b and 3 (Fig. 2B). Incomplete divisions of GCs occur in the region 1, the anterior part of the germarium. Initially, the 16-cell cysts are rounded in shape and multiple cysts can reside next to each other in region 2a of the germarium (Spradling 1993). During these initial stages of GC differentiation, ECs send protrusions around GCs and regulate their differentiation (Schulz et al. 2002, Kirilly et al. 2011, Banisch et al. 2017). As 16-cell cysts transition into region 2b, they flatten to a disc shape and become surrounded by follicle cells. Cysts grow larger and rounder as they reach region 3, where stage 1 of oogenesis begins.

118 Finally, a follicle containing a cyst surrounded by epithelial FCs pinches off from the germarium
119 and transitions through 14 stages of oogenesis to give rise to a mature egg (Fig. 1C, 2B). One of
120 the 16 cells is specified as an oocyte and is destined to become the egg, while the other 15 cells
121 become nurse cells that support oocyte growth by providing mRNAs and proteins. Towards the
122 end of oogenesis during stage 11, nurse cells transfer all their content to the oocyte and
123 subsequently undergo apoptosis (Spradling 1993).

124 To identify gene expression profiles specific to distinct stages of GC differentiation, we sub-
125 clustered the GC I and GC II clusters and obtained 9 clusters (Fig. 2A). To determine the identity
126 of each cluster, we identified cluster specific marker genes (Fig. 2C, Supplemental Table 1),
127 selected markers that are not or are lowly expressed in other cell types (Supplemental Fig. S2),
128 and visualized their expression patterns using hybridization chain reaction (HCR), a highly
129 sensitive fluorescence *in situ* hybridization (FISH) method (Choi et al. 2018). For gene
130 description and brief functional annotation of genes described refer to Supplemental Text.

131

132 The cluster expressing *CG2887*, *CG6628,* and *CG15398* corresponded to the earliest stages of
133 GC differentiation and contained GSCs, CBs, and 2-cell cysts (cluster GSC/CB/2-cc) (Fig. 2B-E,
134 Supplemental Fig. S3A,B). *bam*, a master regulator of GC differentiation onset (Chen and
135 McKearin, 2003, Xie and Spradling, 1998, Xie and Spradling, 2000), was lowly expressed in 2-cell
136 cysts, displayed the strongest expression in 4-cell cysts (cluster 4-cc), and lower levels at later
137 stage cysts (Fig. 2B,C,E,G, Supplemental Fig. S3B,D). *HP6* was expressed in GSCs and early cysts
138 (clusters GSC/CB/2-cc and 4-cc), while *Orc1* showed the strongest expression in 4- and 8-cell
139 cysts (clusters 4-cc and 8-cc) and low expression in GSCs, 2-cell and 16-cell cysts (Fig. 2C,F,
140 Supplemental Fig. S3C). Thus, *HP6* and *Orc1* co-expression in 4-cell cysts demarcated cluster 4-
141 cc. *CG15628* displayed the strongest expression starting from 8-cell cyst stage demarcating
142 cluster 8-cc, and slowly fades after 16-cell cysts reached region 2b (Fig. 2B,C,G, Supplemental
143 Fig. S3D). The next five clusters corresponded to 16-cell cysts at various stages. Strong
144 expression of *CG3691* and *Nlg2* was first detected in 16-cell cysts in region 2a, thus demarcating

145 clusters 16-cc 2a I and 16-cc 2a II (Fig. 2B,C,H,I, Supplemental Fig. S3E,F). The next cluster
146 contained 16-cell cysts from region 2a and 2b (16-cc 2ab). It was marked by the onset of strong
147 *mnd* expression, which was detected in 16-cell cysts in region 2a, and by *CG3961* whose
148 expression extends into 16-cell cysts in region 2b (Fig. 2B,C,H,J, Supplemental Fig. S3E,G).
149 Strong expression of *CG7255* was first detected in 16-cell cysts in region 2b marking cluster 16-
150 cc 2b (Fig. 2B,C,M, Supplemental Fig. S3H). Finally, onset of strong *CG17270* expression in 16-
151 cell cysts in region 3 marked cluster 16-cc 3, and onset *of slif* expression marked stage 2 follicles
152 (cluster St2) (Fig. 2B,C,M,N, Supplemental Fig. S3H,I). *cry* was predominantly expressed in 16-
153 cell cysts in regions 2b and 3 (clusters 16-cc 2b and 16-cc 3) and was downregulated in stage 2
154 follicles (cluster St 2) (Fig. 2B,C,N, Supplemental Fig. S3I). Thus, by exploring the expression of
155 14 marker genes *in situ*, we have assigned all 9 clusters to specific stages of germ cell cyst
156 development. GCs from later stages of oogenesis were not present in our dataset due to their
157 larger size, and thus, incompatibility with our cell suspension preparation and cell capture
158 methods. The clusters were arranged on UMAP in the order of their differentiation status, with
159 undifferentiated cells at the bottom left corner, and differentiated cells on the right side (Fig.
160 2A). This indicates that differentiation is a predominant source of transcriptome variation in our
161 dataset. The data did not allow us to differentiate nurse cells from oocytes by clustering, and it
162 remains unclear if our dataset contains oocytes.

163 We were unable to clearly separate GSCs from CBs and 2-cell cysts by clustering. This is likely
164 due to the high degree of similarity in the gene expression profiles of GSCs, CBs, and 2-cell
165 cysts, the reliance on translational rather than transcriptional control for early GSC
166 differentiation (Slaidina and Lehmann 2014), and the low abundance of these cell types. An
167 observational study using wild-type ovaries estimated that in each germarium, on average
168 there were 2 GSCs, 1.3 CBs, 1.6 2-cell, 0.9 4-cell, 0.8 8-cell cysts, and 8.4 16-cell cysts. Of the 16-
169 cell cysts, 5.5 were found in region 2a, 1.9 in region 2b, and 1 in region 3 (Drummond-Barbosa
170 and Spradling 2001). We used these observational numbers to estimate the coverage obtained
171 for the different cell types (Supplemental Fig. S1B). Coverage varied between cyst stages,
172 suggesting that some cluster population may not be pure. Our analysis revealed distinct

173 transcriptomes for 9 steps of GC differentiation from GSC to stage 2 of oogenesis and identified
174 markers for most (Fig. 2C), suggesting that the transcriptomes identified may correspond to the
175 developmental stages described previously.

176 Gene expression dynamics during GC differentiation.

177 Next, we characterized gene expression dynamics during GC differentiation more closely. 813
178 genes were differentially expressed between GC clusters, indicating that even though the first
179 steps of differentiation rely heavily on translational regulation, there are significant changes in
180 mRNA levels (Slaidina and Lehmann 2014). We visualized the mean expression levels of
181 differentially expressed genes on a heatmap (Fig. 2K, Supplemental Table 1, Supplemental
182 Table 2). 16-cell cyst clusters displayed distinct expression signatures. Differences observed
183 between the five germarial 16-cell cyst states reveal distinct in mRNA expression patterns as
184 16-cell cysts transition from cyst cell division to differentiation. In region 2a 16-cell cysts are
185 enwrapped by ECs , while in region 2b and 3, GCs are surrounded by pre-FCs, which start
186 forming the follicular epithelium. Thus, transcriptional differences may also reflect the changing
187 signals GC cysts receive from various bordering somatic cell types.

188 We explored the functional predictions for these genes using Gene List Annotation for
189 *Drosophila* (GLAD) (Hu et al. 2015) (Supplemental Fig. S4A, Supplemental Table 3). GC stages in
190 contact with ECs (up to 16-cell cysts in region 2a) were enriched for transmembrane proteins,
191 some of which may be involved in signaling between GCs and ECs, and others in formation of
192 the fusome, which is composed of ER derived vesicles (Huynh 2006). Furthermore, the
193 GSCs/CB/2-cc cluster was enriched for chaperones and heat shock proteins, which have a
194 protective function. Consistent with divisions of early cysts, GSC/CB/2-cc and 4-cc clusters were
195 enriched for cytoskeletal proteins, in particular, those involved in cytokinesis. In contrast, most
196 late stage cysts were enriched for mitochondria related genes, which coincide with rapid
197 expansion of the mitochondrial pool and their selection based on fitness (Hill et al. 2014, Lieber
198 et al. 2019). Finally, transcription factors and DNA-binding proteins were significantly enriched
199 at multiple stages over the course of GC differentiation, and we visualized their expression in a

200 heatmap (Fig. 2L). A few were expressed in GSC/CB/2-cc cluster, including, *CG15398* (Fig. 1D,
201 Supplemental Fig. S3A) and *chinmo*, which regulates cyst stem cell renewal in *Drosophila* testis
202 (Flaherty et al. 2010), suggesting that it may have a conserved role in stem cell renewal across
203 many tissues. Five Notch responsive *Enhancer of split Complex* (*E(spl)-C*) transcription factors
204 were expressed in 4-cc, 8-cc and 16-cc 2a I clusters, suggesting that Notch signaling pathway
205 may regulate early steps in GC differentiation.

206 Altogether, we have identified specific transcriptional signatures for 9 distinct steps of GC
207 differentiation. Further analyses of gene expression dynamics and genetic requirements
208 between the clusters may reveal novel aspects of GC differentiation regulation.

209 *Drosophila* germaria contain three EC subtypes.

210 Next, we aimed to identify transcriptional signatures for somatic cells of the germarium. There
211 are five major somatic cell types in germaria. At the most anterior tip of the germarium reside
212 the somatic terminal filament (TF) and cap cells (CC) of the GSC niche (Sahut-Barnola et al.,
213 1995) (Fig. 3B). Directly posterior to them in regions 1 and 2a are ECs, which extend protrusions
214 that enwrap GCs (Schulz et al., 2002). Posterior to the ECs at the region 2a and 2b boundary are
215 FSCs, which divide asymmetrically to give rise to a transit amplifying FC population called pre-
216 follicle cells (pre-FC) (Margolis and Spradling, 1995). Pre-FCs reside in regions 2b and 3 where
217 they proliferate and further differentiate into polar, stalk and epithelial FC lineages.

218 To assign transcriptomes to individual cell types, we subclustered the 'germarium soma' cluster
219 and initially obtained seven clusters. A small number of cells located separately on UMAP and
220 were included in one of the larger clusters by the clustering algorithm. These cells expressed *en*
221 and *Lmx1a,* well known TF and CC markers (Bolívar et al. 2006, Allbee et al. 2018) (Fig. 3A-D,
222 Supplemental Fig. S1E, S3J). Therefore, we manually annotated this cell population as TF/CCs
223 (see Supplemental Text). Each ovariole contains only about 8 TFs and 6 CC (Godt and Laski,
224 1995, Xie and Spradling, 2000). In total, our dataset contained only 10 such cells, and thus we
225 were unable to distinguish TFs from CCs. We computed markers for all 8 clusters (Supplemental
226 Table 1). Of the remaining seven clusters, four expressed *Fas3*, *cas*, or both, suggesting they

227 represent FC subtypes (Fig. 3L) (Zhang and Kalderon, 2001, Chang et al., 2013), and the
228 remaining three expressed high levels of *fax* and *Wnt4* (Decotto and Spradling 2005, Mottier-
229 Pavie et al. 2016), suggesting that they correspond to three distinct subtypes of ECs (Fig. 3A-D,
230 Supplemental Fig. S3J). EC subtypes may correspond to ECs that are functionally distinct, or that
231 are in three distinct anatomical positions. Such subdivision of ECs along the anterior posterior
232 axis in the germarium has been suggested on the basis of morphological differences and limited
233 functional analyses (Banisch et al., 2017; Eliazer et al., 2014; Rojas-Rios et al., 2012). To
234 distinguish between these possibilities, we explored marker gene expression of each cluster.
235 *hh*, *Wnt4* and *ptc* are expressed in an anterior to posterior gradient, and their expression
236 differed between EC clusters; suggesting the three clusters might correspond to distinct
237 anatomical positions - anterior (aEC), central (cEC), and posterior (pEC) (Fig. 3A-E). Moreover,
238 *Wnt6* functions in anterior ECs (Wang and Page-McCaw 2018), and was enriched in the aEC
239 compared to other EC clusters (Fig. 3C). We selected a few EC subtype markers, and assessed
240 their mRNA expression in the ovary. *NetA* and *Notum* were specifically expressed in the aEC
241 cluster (Fig. 3C), and, indeed, labeled anterior ECs in region 1 (Fig. 3F, Supplemental Fig. S3L).
242 *CG31431* and *CG10073* were expressed in the cEC cluster (Fig. 3C), and the strongest expression
243 of their mRNAs was detected in the central region of the germarium, in close contact to 8- and
244 16-cell cysts (Fig. 3G,H, Supplemental Fig. S3N,O). We used a combination of two markers to
245 identify the remaining EC population. *wun2* is predominantly expressed in cEC and pEC clusters,
246 while *AdamTS-A* is expressed in the pEC and FC lineage (Fig. 3C); therefore, cells co-expressing
247 *wun2* and *AdamTS-A* correspond to pEC. We observed that *wun2* is predominantly expressed in
248 ECs in Region 2a, and *AdamTS-A* in the early FC lineage, and expression of both genes
249 overlapped near region 2a and 2b boundary (Fig. 3I, Supplemental Fig. S3M, arrowheads).
250 These results suggest that the aEC, cEC, and pEC clusters correspond to the most anterior,
251 central and posterior ECs. EC transcriptomes are aligned from anterior to posterior as indicated
252 by a pink arrow in the UMAP plot (Fig. 3A, pink arrow).
253

254 Two additional pEC markers, *kar* and *peb*, were expressed in a fraction of pEC that lacked
255 expression of *wun2* and *AdamTS-A* (Fig. 3A, dotted line in magenta, 3C, Supplemental Fig. S1E).

256 In the germarium, *kar* and *peb* displayed sparse weak expression throughout the germarium
257 (Fig. 3H,K, Supplemental Fig. S3O,P), suggesting that there might be additional complexity
258 within the EC population that has not been revealed by clustering. Indeed, it has been
259 suggested that ECs can migrate towards the anterior to possibly replace lost ECs (Reilein et al.,
260 2017).

261 To test whether we captured all EC subtypes, we labeled *laza*, which is expressed in aEC and
262 cEC together with *wun2* and *AdamTS-A*, and observed that all regions of the germarium where
263 labeled, except the most anterior tip (TF/CC) (Fig. 3I, Supplemental Fig. S3M). Thus, we have,
264 indeed, captured all EC subtypes.

265 Altogether, we have identified three transcriptionally distinct EC subtypes arranged along the
266 anterior-to-posterior axis of the germarium. The number of cells recovered corresponded well
267 to previous observations (Supplemental Fig. 1B). To explore the potential functional differences
268 between EC subtypes in our dataset, we compared their transcriptomes. 405 genes were
269 differentially expressed between EC clusters. We visualized their expression in a heatmap
270 together with other somatic cell markers (Fig. 3M) and performed GLAD analyses on EC marker
271 genes (Fig. S4B, Supplemental Table 3). Consistent with their role as GC differentiation
272 regulators, ECs were strongly enriched for gene classes related to cell-cell communication, like,
273 signaling pathway components, secreted proteins, receptors, transmembrane proteins, and
274 extracellular matrix (matrisome) proteins. Moreover, to facilitate discovery of transcription
275 factors regulating distinct EC fates, we have visualized expression of differentially expressed
276 transcription factors (Supplemental Fig. S4C).

277 Because of their anatomical position, the three EC populations are in contact with GCs at
278 different states of differentiation, suggesting that each EC subtype might send a distinct set of
279 signals to regulate different stages of GC self-renewal or differentiation. To identify such
280 potential regulators, we visualized the expression of all secreted proteins and ECM components
281 that are differentially expressed among ECs subtypes (Fig. 3J).

282 *Fs* (*Follistatin*), an inhibitor of activin and TGF-beta signaling (Bickel et al. 2008), was expressed
283 specifically in cECs. Since the signaling of TGF-beta homolog *dpp* promotes GSC self-renewal
284 and impedes GC differentiation, it is plausible that *Fs* expression in the cECs may facilitate GC
285 differentiation by suppressing the *dpp* signal emanating from the GSC niche. Thus, *Fs*
286 expression in cEC may provide an additional layer of BMP signaling regulation in the germarium
287 (Song et al., 2004) (Supplemental Fig. S4D).

288 aECs expressed Wnt signaling pathway components and regulators, including ligands *Wnt4* and
289 *Wnt6*, and *Notum,* which regulates Wnt ligand activity (Kakugawa et al. 2015). aECs expressed
290 also EGF and FGF ligands *vn* and *ths.* Indeed, we observed higher levels of pERK, a EGFR and
291 FGFR pathway readout, in the anterior and central part of the germarium (region 1 and 2), than
292 in region 3 (Supplemental Fig. S4E). Insulin-like peptide *Ilp6* was specifically expressed in cEC,
293 and two regulators of insulin/IGF signaling *Idgf1* and *Idgf3* were expressed in aEC and cEC. Thus,
294 a number of signaling pathway activators and regulators were expressed in distinct regions of
295 the germarium and may regulate GC differentiation. Moreover, ECs may also signal to each
296 other and to additional somatic cells of the germarium, like, TFs and CCs, or FSCs. Indeed, it has
297 been suggested that ECs regulate FSC maintenance (Sahai-Hernandez and Nystul, 2013; Vied et
298 al., 2012). cECs expressed also a number of secreted peptidases (*Cys*, *CtsB1*, *Cp1*), which may be
299 involved in signaling molecule activation or degradation. Moreover, peptidase inhibitors *nec*,
300 *CG17739* and *Kal1* were expressed in adjacent EC clusters and, thus may refine the peptidase
301 activity domain. pECs were enriched for genes required for collagen production and secretion
302 (*PH4alphaEFB*, *Plod*), while aEC and cEC expressed matrix metalloprotease (*Mmp1*) and its
303 inhibitor (*Timp*), which regulate ECM structure, suggesting that there is a distinct ECM structure
304 and composition in different parts of the germarium. Moreover, a number of adhesion
305 molecules and axon guidance molecules (*NetA*, *Ten-a*, *Sema1a*) where differentially expressed
306 between the EC subtypes. Altogether, these results indicate that each EC subgroup secretes a
307 different set of molecules, which may create distinct microenvironments for GC development
308 and differentiation along the anterior-posterior axis of the germarium. Since signaling between
309 ECs and GCs relies at least in part on large surface contacts between GCs and EC protrusions

310 (Schulz et al. 2002; Kirilly et al., 2011; Banisch et al., 2017), we sought to search for physical
311 interactor pairs (such as transmembrane-transmembrane and secreted-transmembrane) that
312 may mediate signaling between GCs and ECs (Guruharsha et al. 2011). We selected secreted
313 ligands, extracellular matrix proteins (ECM), and transmembrane proteins enriched in ECs (158
314 genes) or GC clusters adjacent to ECs (GSC/CB/2-cc, 4-cc, 8-cc, 16-cc 2a I, 16-cc 2a II, and 16-cc
315 2ab) (200 genes). 49 genes (16 expressed in GCs, 23 in ECs, and 10 in both) formed 119
316 predicted protein: protein interaction pairs between transmembrane/secreted proteins in GCs
317 and ECs (Supplemental Table 4). In 30 pairs, both genes had GO terms associated to plasma
318 membrane or extracellular space, and in 65 pairs at least one of the genes was uncharacterized
319 (in 17 both were uncharacterized). Further analyses of the physical interaction pairs may reveal
320 novel mechanisms for GC differentiation regulation by ECs.

321 Altogether, our data together with a recent reports by Rust et al. 2020, Tu et al 2020 and Shi et
322 al. identifies multiple EC subtypes and suggests functional specialization among ECs. Subtype
323 specific markers, transcriptomes, and potential signaling molecules and interactors provide a
324 resource for the functional characterization of EC subtypes and may reveal novel signaling
325 mechanisms between GCs and soma.

326 Identification of early FC lineages

327 We next sought to characterize the FC lineages present in the germarial somatic cell clusters.
328 One cluster was rather distant from other clusters and expressed polar cell markers *upd1* and
329 *upd3* (Fig. 3A-C). Visualization of two newly identified markers *CG34377* and *Hk*, confirmed that
330 the cluster corresponded to polar cells (cluster polar) (Fig. 3C,O, Supplemental Fig. 3R).

331 Stalk cells express *LamC* and *cas* (Song and Xie 2003; Chang et al. 2013). Two adjacent clusters
332 expressed both markers, albeit at different levels (Fig. 3A,C). To determine the cluster identity,
333 we searched for genes that were enriched and differentially expressed between the clusters.
334 *sim* was expressed in both, while *CG10311* and *AdamTS-A* were each enriched in one of the
335 clusters (Fig. 3C). *sim* was expressed in FCs at region 3 and stage 2 follicle interface, and in stalk

336 cells (Fig. 3N, Supplemental Fig. S3Q), suggesting that the two clusters correspond to stalk cells
337 and FCs differentiating into stalk cells (pre-stalk). *CG10311* was strongly expressed in stalk and
338 weakly expressed in pre-stalk cells, while *AdamTS-A* was expressed in pre-stalk (and other pre-
339 FCs), but absent from stalk cells, indicating that the *AdamTS-A* expressing cluster corresponds
340 to the pre-stalk cells, and *CG10311* expressing cluster corresponds to the stalk cells (Fig. 3C,N,
341 Supplemental Fig. S3Q).

342 Finally, the remaining cluster expressed high levels of *AdamTS-A* and *CG1136*, which are both
343 expressed in pre-FCs in germarium regions 2b and 3 (Fig. 3K,L). Therefore, we concluded that
344 this cluster contains FSCs and pre-FCs (FSC/pre-FC). Consistently, this cluster expressed also
345 high levels of the pre-FC marker *eya* (Fig. 3C) (Bai and Montell 2002). Our clustering analyses
346 did not reveal a clear FSC population, likely because the low numbers of FSCs per germarium
347 resulted in too few FSC transcriptomes present for our analyses (Fadiga and Nystul 2019,
348 Reilein et al. 2017), or their transcription profiles are highly similar to pre-FCs (Nystul and
349 Spradling 2010, Hartman et al. 2015, Reilein et al., 2017, Rust and Nystul, 2020).

350 Aiming to identity the FSCs and their markers, we performed pseudotime analyses on follicle
351 cell lineages from germarium soma clusters (FSC/pre-FC, pre-stalk, stalk, polar), and early
352 epithelial FC (St 2-4I, St2-4 II) using Monocle, an excellent tool to align cells on a “pseudotime"
353 axis and study gene expression dynamics over the course of differentiation (Qiu et al. 2017).
354 These analyses revealed three branches of FC differentiation, one corresponding to epithelial
355 FCs, one to polar cells and the third to stalk cells, where FCs transition through a pre-stalk state
356 before acquiring stalk cell identity (Fig. S5A,B) (Assa-Kunik et al., 2007). *eya* was downregulated
357 as FCs committed to polar and stalk cell fate, and remained highly expressed in epithelial FCs
358 (Fig. S3C). In contrast, *CG10311* was upregulated and *AdamTS-A* was downregulated in stalk
359 cells over time, and *upd1* was upregulated in polar cells (Fig. S3D-F). Thus, our Monocle analysis
360 agrees well with literature reports (Tworoger et al. 1999, Beccari et al. 2002) and with our
361 cluster assignment to cell types. We, therefore, used the results of this analysis for
362 identification of putative FSCs and their markers.

363 Monocle analyses suggested two putative markers for FSCs, *if* and *CG6044* (Fig. S5G-I), and we
364 visualized their expression using HCR. To discern the precise location of the region 2a/2b
365 boundary where FSCs are positioned, we labeled *Fas3* mRNA, which labels the follicle cell
366 lineage, and Dapi, which labels cell nuclei. We found that *if* was expressed in the region where
367 FSCs reside, however, its expression extended to pre-FCs and ECs, where it was weakly
368 expressed (Fig. S5J). *CG6044* was predominantly expressed in the pre-FCs, with weak expression
369 at the germarium region 2a and 2b boundary where FSCs reside (Fig. S5K). Therefore, neither *if*
370 nor *CG6044* are FSC specific markers. We were unable to identify FSC specific markers, likely
371 because the FSC gene expression profile is highly similar to the pre-FC gene expression profiles
372 (Nystul and Spradling 2010; Hartman et al. 2015; Reilein et al., 2017), and therefore they are
373 indistinguishable by bioinformatics in our dataset. Increasing the number of cells of interest
374 might increase the statistical power to alleviate these limitations. Rust *et al.* used a similar
375 approach to identify FSCs and their markers. While their markers are, indeed, enriched in our
376 putative FSC population, their mRNA expression extends to ECs and pre-FCs similarly to the
377 genes we identified as marking FSC populations (Supplemental Text, Supplemental Fig. S10C).

378 399 genes were differentially expressed between the FC clusters, and we visualized their
379 expression in a heatmap (Fig. 3M). The polar cell cluster displayed the most distinct signatures,
380 while pre-stalk and stalk cells shared a fraction of their marker genes. We performed GLAD
381 analyses on all marker genes (Supplemental Fig. S4B). Major signaling pathway components
382 were enriched in FC lineages, consistent with high signaling activity during FC lineage
383 specification. FC lineage clusters were also enriched for cytoskeletal genes. The FSC/pre-FSC
384 cluster was particularly enriched for cell division related cytoskeletal proteins, consistent with
385 active proliferation. A number of TF/CC specific genes were also enriched in stalk cells. Both cell
386 types are post-mitotic and share a similar flattened shape, suggesting that many of the shared
387 genes may comprise cell shape regulators. Moreover, we have visualized expression of
388 transcription factors that are differentially expressed in follicle cell lineages (Supplemental Fig.
389 S4C). Further studies will determine whether they are required for acquisition of distinct FC
390 fates. Altogether, we determined FC lineage gene expression profiles and markers. Exploration

391 of the gene expression dynamics as FSCs transition into specialized FC fates may reveal novel
392 regulatory mechanisms for specification of FC lineages.

393 ## Identification of epithelial FC subtypes.

394 Epithelial follicle cells together with GCs transition through 14 stages of oogenesis and are the
395 most abundant ovarian cell type. FC and GC progression through oogenesis is highly
396 coordinated, since FCs play an essential role in egg development by determining the body axes
397 of the future embryo and producing essential yolk proteins (Wu et al. 2008). Determining the
398 transcriptional differences between FC subtypes will strengthen our understanding of how this
399 coordination is achieved. Because we manually removed late stage follicles during dissections
400 (see Materials and Methods), we expected to only identify FC transcriptomes from early and
401 mid-stage follicles. Indeed, the stage 9/10A FC marker *Yp2* was well expressed, but the stage
402 10B marker *Cp36* expression was barely detectable in our dataset (Supplemental Fig. S6A)
403 (Tootle et al. 2011), indicating that our dataset contains FC transcriptomes up to stage 9/10A.
404 Therefore, we focused on FC development up to that stage.

405 Follicle cells progress from mitotic (stage 1-6) to transitioning (stage 7) to endocycling states
406 (stage 7 - 9) (Deng et al. 2001) (Fig. 4D). Starting from stage 5, FCs differentiate into multiple
407 subtypes determined by external cues, which largely depend on the anatomical position within
408 the follicle (Wu et al. 2008; Duhart et al., 2017). FCs at the anterior and posterior tip of the
409 follicle receive JAK/STAT signal (*upd1* and *upd3*) from polar cells**,** and assume terminal fate,
410 while the rest become MainBody (MB) follicle cells (MBFC) (Xi et al. 2003). Anterior terminal
411 (AT) follicle cells (ATFC) assume three distinct fates determined by the distance from the tip -
412 border cells, stretched cells, and centripetal cells (Xi et al. 2003). Posterior terminal cells receive
413 an additional EGF signal (*grk*) emanating from the oocyte nucleus starting from stage 6 and
414 assume the posterior terminal (PT) follicle cell (PTFC) fate (González-Reyes et al. 1995). Initially,
415 epithelial FCs have cuboidal shape, and, at stage 9, the PTFCs transition into columnar shape
416 FCs. Simultaneously, border cells together with anterior polar cells start migrating posteriorly
417 between the nurse cells. Stretched cells start producing a TGF-beta ligand (*dpp*), and in

418 response to it flatten in a wave from the anterior to posterior covering the rapidly growing
419 nurse cells (Brigaud et al. 2015). The MBFCs cover the growing oocyte and assume columnar
420 shape as well. At stage 9/10A, the oocyte nucleus migrates to the anterior dorsal corner of the
421 oocyte, and by activating EGFR signaling specifies the dorsal FCs (Spradling 1993, González-
422 Reyes et al. 1995). Centripetal cells, which initially are located between stretched cells and
423 MBFCs, start migrating between the nurse cells and oocyte at stage 10B (Wu et al. 2008).
424 Further follicle cell morphogenesis events occur beyond this stage, and each follicle cell subtype
425 has a specific role in forming various eggshell structures (Duhart et al. 2017).

426 To find gene expression profiles of epithelial follicle cell subpopulations, we performed in-depth
427 analyses on FC I, II and III clusters. We first assessed expression of known markers to assign
428 transcriptomes to distinct stages of oogenesis and follicle cell subtypes.

429 For staging, we assessed the expression of *ct*, *stg*, *CycB* and *peb*. *ct* and *CycB* are expressed in
430 mitotic FCs up to stage 6, while *stg* expression extends into stage 7 (Fig. 4D,G, Supplemental
431 Fig. S6A, S7B) (Jackson and Blochlinger 1997; Sun and Deng 2005; Deng et al. 2001).
432 Transcription factor *peb* (also called *Hindsight*) is expressed sporadically starting at stage 6 and
433 its expression is strongly upregulated during stage 7 and beyond (Supplemental Fig. S6A) (Sun
434 and Deng 2007). As mentioned above, *Yp2* marks stage 9 follicles (Fig. 4D,N, Supplemental Fig.
435 S6A, S7H) (Tootle et al. 2011). Therefore, the end of stage 6 is marked by downregulation of
436 *CycB* and *ct*, stage 7 is marked by overlapping expression of *stg* and *peb* (Supplemental Fig.
437 S6A), *peb* expression in absence of *stg* marks FCs starting from stage 8, and *Yp2* marks FCs
438 starting from stage 9. Therefore, in the UMAP plot (Fig. 4A), transcriptomes of early stage FCs
439 are located in the bottom left corner and transition in an arch towards the right where FCs
440 differentiate into multiple subtypes.

441 Next, we sought to assess whether the anatomical position correlated with specific UMAP
442 regions. We assessed expression patterns of a number of MB and terminal follicle cell markers.
443 *mirr* marks MBFCs, and was expressed in a band in the middle of the UMAP (Fig. 4A,I,K,
444 Supplemental Fig. S6A, S7D,F) (Jordan et al. 2000). *mid*, *H15* and *pnt* are expressed at the

445 posterior terminus at stage 6 and onwards, and we observed their expression in a band on the
446 bottom of the UMAP (Fig. 4A,I,K Supplemental Fig. S6A, S7D,F) (Morimoto et al. 1996, Lomas et
447 al. 2013). Following dorsal migration of the oocyte nucleus at stage 9, *pnt*, but not *mid* or *H15,*
448 expression is induced in anterior dorsal cells (Morimoto et al. 1996, Lomas et al. 2016). In our
449 dataset, *pnt* was always co-expressed with *mid* and *H15*, indicating that the anterior dorsal
450 stage 9/10A cells are absent in our dataset. *dpp* is expressed in stretched cells, and we
451 observed its expression in the top right corner of the UMAP (Supplemental Fig. S6A) (Xi et al.
452 2003). The border cell and centripetal cell marker *slbo* was detected in a very small number of
453 cells adjacent to *dpp* expressing cells (Supplemental Fig. S6A) (Murphy et al. 1995). Therefore,
454 AT follicle cells are on the top, MB in the middle and PT on the bottom of the UMAP plot (Fig.
455 4A), indicating that the transcriptome variation in FCs is driven by both - the anatomic position
456 in the follicle and the oogenesis stage.

457 To obtain precise subdivision of FCs by stage and subtype, we subclustered the transcriptomes
458 into 15 clusters, and manually split 2 clusters (Supplemental Text) (Fig. 4A). We identified
459 specific markers for each cluster, and visualized their expression along with a few previously
460 described markers (Fig. 4B,E-N, Supplemental Fig. S7A-H).

461 *CG1136* was predominantly expressed in FCs of very early follicles up to stage 4 [clusters St2-4 I,
462 St2-4 II] (Fig. 4B,D,E, Supplemental Fig. S6A, S7A). *CG15546* expression was patchy in stage 2-4
463 follicles [clusters St2-4 I, St2-4 II, PT St4-5], and slightly enriched at the termini prior to
464 enriching to the PT at stage 6 [cluster PT St6] (Fig. 4B,D,G, Supplemental Fig. S6A, S7B), while
465 *CG31808* had weak patchy expression in stage 2-4 follicles and restricted predominantly to
466 ATFCs at stage 5-6 [clusters St2-4 I, St2-4 II, PT St4-5, and AT St5-6] (Fig. 4B,D,F, Supplemental
467 Fig. S6A). *ct* was expressed in FCs up to stage 6 as previously reported [clusters St2-4 I, St2-4 II,
468 PT St4-5, MB St5-6, MB St6, AT St5-6, PT St6] (Fig. 4B,D,G, Supplemental Fig. S6A, S7B) (Jackson
469 and Blochlinger 1997). *Asph* was ubiquitously expressed starting from stage 5 and enriched at
470 both termini by stage 7 [clusters MB St5-6, MB St6, AT St5-6, PT St6, AT St7, PT St7] (Fig.
471 4B,D,E,G, Supplemental Fig. S6A, S7A,B). MBFCs were labeled by patchy expression of *CG10924*
472 at stage 6-7 [clusters MB 5-6, MB St6, MB St7], and the previously described transcription factor

473 *mirr* from stage 6 onwards [clusters MB St7, MB St8, MB St8-9, MB St9] (Fig. 4B,D,H,I,K,
474 Supplemental Fig. S6A, S7C,D,F) (Jordan et al. 2000).

475 We visualized expression of four genes expressed in AT. *Spn42Dd* was predominantly expressed
476 in the ATFCs at stages 5-7 [clusters AT St5-6, AT St7] (Fig. 4B,D,H, Supplemental Fig. S6A, S7C).
477 *magu* was expressed in AT and lowly at PT starting stage 6, peaking at stage 8, and declining at
478 stage 9 [clusters AT St7, AT St8-9] (Fig. 4B,D,I, Supplementals Fig. S6A, S7D). *Amph* was
479 expressed in ATFCs at stage 8 [cluster AT St8-9], and its expression was undetectable in
480 stretched cells at stage 9 (Fig. 4B,D,J, Supplemental Fig. S6A, S7E), while *CG5002* was expressed
481 in stretched cells at stage 9 [cluster AT St8-9] (Fig. 4B,D,J, Supplemental Fig. S46A, S7E). *ImpL2*
482 expression started in ATFCs at stage 8 and reached far posterior in MBFCs but was excluded
483 from the very anterior tip [clusters AT St8-9, MB St8-9, MB St9]. In stage 9, *ImpL2* was
484 expressed in flattened stretched cells and border cells (Fig. 4B,D,E,J,L,N, Supplemental Fig. S6A,
485 S7A,E,G-H).

486 PTFCs were faintly labeled by *laza* and *mid* at early stages, reaching maximum at stage 8
487 [clusters PT St7, PT St8 I, PT St8 II] (Fig. 4B,D,I,K, Supplemental Fig. S6A, S7D,F). *CG8303* was
488 enriched in PTFCs at stages 8 and 9 [clusters PT St8 I, PT St8 II, PT St9], *Yp2* and *bond* at stage 9
489 [cluster PT St 9] but all three were expressed in MBFCs as well [clusters MB St9, and MB St8-9
490 for CG8303 only] (Fig. 4B,D,L,M,N Supplemental Fig. S6A, S7G,H). Stage 8 PT FCs were divided in
491 two sub-populations: cluster PT St8 II corresponded to the most posterior tip and expressed
492 *H15* along with *laza* and *mid*, and cluster PT St8 I corresponded to a band of cells adjacent to PT
493 St8 II cells, which expressed *laza* and *mid* and only low levels of *H15* (Fig. 4B,D,I,K, Supplemental
494 Fig. S6A, S7D,F).

495 Characterization of FC transcriptional signatures

496 518 genes were differentially expressed between FC subtypes, and we visualized their
497 expression in a heatmap (Fig. 4C). We observed temporal signatures (stage 2-6, and stage 8-9)
498 and cell fate dependent (ATFC, MBFC, PTFC) signatures. The gene expression signature of each

499 cluster was composed of blending of both the temporal and cell fate signatures, especially
500 during earlier stages of oogenesis.

501 To gain insights of functional differences between FC subtypes, we explored gene class
502 enrichment in the signature of each cluster (Supplemental Fig. S6B-D, Supplemental Table 3)
503 (Hu et al. 2015). Transcription factors/DNA-binding proteins were enriched in multiple clusters,
504 highlighting the role of transcriptional regulation in follicle cell subtype diversification.
505 Cytoskeletal proteins were enriched in multiple early stage follicle cell subtypes and in ATFCs at
506 later stages. The sets of genes, however, were distinct. The early proliferative FCs were
507 enriched for cell division associated cytoskeletal proteins, while stage 7 and 8-9 ATFCs were
508 predominantly enriched for actin binding proteins and actin regulators, many of which are
509 known to be involved in cell morphogenesis events such as axon targeting. In FC cells, they are
510 likely involved in stretched cell flattening.

511 Transporters, in particular, vacuolar ATPase (V-ATPase) subunits were enriched in late stage
512 follicle cells, especially in cluster ATFCs St8-9. V-ATPase is a large multimolecular complex that
513 uses the energy of ATP hydrolysis to create proton gradients, for example, to lower the pH of
514 particular cell compartments, like, lysosomes or endosomes (Nelson et al. 2000). V-ATPases
515 locate to the stretch cell plasma membrane at stage 13 of oogenesis to acidify nurse cells and
516 facilitate their death and engulfment by stretch cells (Timmons et al. 2016). It is plausible that
517 FCs begin upregulating V-ATPase subunit genes as early as stage 8-9 in preparation for this
518 acidification. Indeed, the first signs of nurse cell death can be observed already at stage 10
519 (Cooley et al. 1992), suggesting that V-ATPases may start executing their function well prior to
520 stage 13.

521 Furthermore, we visualized expression of transcription factors that are differentially expressed
522 among epithelial FC subtypes (Supplemental Fig. S6E). In addition, numerous other gene classes
523 were enriched in individual clusters, and functional studies may reveal new insight into specific
524 roles of follicle cell subtypes. Moreover, in-depth studies of differentially expressed

525 transcription factors will help understand the transcriptional networks governing complex FC
526 subtype specification and morphogenesis.

527 Discussion

528 We generated a single cell atlas of the stem cell compartment and early differentiating egg
529 chambers of adult ovaries of *Drosophila melanogaster*. We characterized cell type-specific
530 transcriptional signatures and identified novel markers. In only few cases does a single maker
531 gene uniquely identify specific cell type, but in combination and by intensity of gene expression,
532 these genes can be used as cell type markers that distinguish between related cell types and
533 developmental transitions. With these transcriptional profiles we generated functional
534 predictions for 34 cell types and subtypes - 9 states of GC differentiation, and 25 somatic cell
535 types including GSC niche cells, 3 EC subtypes, FSC/pre-FCs, 3 clusters corresponding to polar
536 and stalk cell lineages, and 17 epithelial FC subtypes. This extensive annotation will bolster
537 future studies, for example, by capitalizing on cluster specific markers to develop cell type
538 specific genetic tools and for the analysis of stage and cell type specific functions.

539 We were unable to distinguish GSCs and FSCs from their daughters by clustering. This is likely
540 due to the relatively low numbers of stem cells in our dataset, but it also reflects the high
541 similarity between the transcriptomes of stem cells and their daughters (Kai et al. 2005; Slaidina
542 and Lehmann 2014; Nystul and Spradling 2010; Hartman et al. 2015; Reilein et al., 2017; Rust
543 and Nystul, 2020). This predurance of the stem cell transcriptome might have a functional
544 relevance. Indeed, CBs and germline cysts can de-differentiate and compete with GSCs for
545 niche occupancy (Xie and Spradling 1998; Liu et al. 2015). Likewise, FSC daughters migrate
546 across the germarium and possibly compete for niche occupancy with other FSCs (Nystul and
547 Spradling 2007, Reilein et al., 2017). Thus, stem cell daughters initially retain the ability to
548 revert to a stem cell state, possibly because they do not extensively remodel their
549 transcriptome immediately after the asymmetric division.

550 ECs have a dual role, they promote GSC self-renewal and GC differentiation forming a domain
551 termed differentiation niche (Schulz et al., 2002, Kirilly et al. 2011, Rojas-Rios et al., 2012, Liu et
552 al., 2015, Wang and Page-McCaw 2018). We and Rust *et al*. identified three EC subtypes -
553 anterior, central, and posterior ECs. Similarly, Shi et al. and Tu et al. report multiple EC
554 subtypes. Each EC subtype interacts with GCs of a particular differentiation state, and likely
555 sends and receives distinct signals (Tu et al., 2020). Likewise, EC morphologies differ between
556 locations; cEC and pEC protrusions are longer than aEC protrusions, as they interact with
557 increasingly larger germline cysts (Banisch et al. 2017). We uncovered that a number of
558 secreted proteins, adhesion molecules and ECM components are differentially expressed
559 between EC subtypes, suggesting that each subtype creates a distinct microenvironment. Thus,
560 as GCs progress through differentiation and move posteriorly along the germarium, their
561 immediate microenvironment changes. These observations open the possibility that the spatial
562 organization of distinct EC microenvironments supports progressive GC differentiation, and that
563 maturing GC may feed back on their microenvironment to define and stabilize its pattern.

564 Highly granular clustering of GC transcriptomes and precise cluster identity assignments
565 allowed us to identify transcription factors that are dynamically expressed over the course of
566 differentiation. We uncovered that a number of Notch signaling responsive transcription factors
567 (*Enhancer of split Complex*) are enriched in 4- and 8-cell cysts raising a possibility that Notch
568 signaling regulates early steps of GC differentiation.

569 Two recent studies have generated similar adult ovary atlases (Rust et al. 2020; Jevitt et al.
570 2020). Two additional studies have focused mainly on escort cells (Shi et al. 2020; Tu et al.
571 2020). Each study had a distinct focus and approach (for a detailed comparative analysis of
572 these studies and discussion refer to Supplemental Text and Supplemental Figures S8 -S10).
573 Overall, our, Rust et al, and Jevitt et al. studies produced similar results (Supplemental Fig. S8,
574 S9), despite technical differences in fly lines, sample preparation and analytical methods. The
575 majority of markers identified showed the expected expression in our dataset (Supplemental
576 Fig. 10, Supplemental Text). The exact cluster boundaries occasionally differ between the
577 studies, for example, for EC subtypes. Shi et al. and Tu et al. focused more specifically on ECs to

578 understand their functional diversity. Our study provides a more granular sub-clustering of
579 ovarian cell types, and precise cluster mapping to cell types and differentiation stages by direct
580 visualization of mRNAs *in situ*. Thus, we deliver a comprehensive resource of gene expression
581 profiles and markers for each cluster, and provide gene class annotations for transcriptional
582 signatures and functional predictions.

583 Despite being one of the most extensively studied adult organs, our analysis of the *Drosophila*
584 ovary reveals higher cell type diversity than previously anticipated. These findings suggest that
585 numerous, yet unidentified, cell subpopulations with distinct functions exist even in the most
586 thoroughly studied organs. The ongoing Human Cell Atlas and similar projects in model
587 organisms will start revealing this complexity, while focused studies are needed to uncover the
588 interplay between the subpopulations and their functions, finally allowing us to fully
589 comprehend organ function in homeostasis and disease.

590 # Materials and Methods

591 For complete, detailed description of materials and all methods including those summarized
592 below see Supplemental Text.

593 ## Ovary dissociation for scRNA-seq

594 For each sample, 30-40 pairs of adult ovaries were dissected one by one in ice-cold DPBS, no
595 calcium, no magnesium (Thermo Fisher Scientific #14190136). Using forceps and a dissection
596 needle, anterior tip of the ovary was cut and transferred to a new well. We used two
597 dissociation methods. Method A: Dissociation by Type I Collagenase and Trypsin and FACS for
598 purification. Method B: Dissociation by Type I Collagenase and elastase and filtration for
599 purification.

600 ## scRNA-seq data analysis

601 We used Seurat v3 for sc-RNA seq analyses following standard guidelines for data processing,
602 identification of variable genes, dimensionality reduction and cell clustering. For coarse
603 clustering, we used resolution parameter 0.2. For step wise clustering, we selected cell groups
604 from initial coarse clustering - GCs, follicle cells or somatic cells of germaria. The data was
605 scaled, new PCAs were computed, and new UMAP plots were generated, followed by clustering
606 with higher resolution parameters (GC - 0.6, somatic cells of germarium - 1, and FCs - 0.6). For
607 marker identification we used Wilcox text and visualized marker expression in heatmaps using
608 *pheatmap* package.

609 We used the GLAD online tool to align marker genes for each cluster with particular gene
610 categories, and assessed significant enrichment in each cluster's transcriptional signature. For
611 physical interaction analyses, we selected genes from following GLAD categories: GPCRs, Ion
612 channels, Matrisome (ECM), Receptors, Secreted proteins, Transmembrane proteins, and
613 Transporters (Hu et al. 2015) from EC cluster markers, and select GC cluster (GSC/CB/2-cc, 4-cc,
614 8-cc, 16-cc 2a I, 16-cc 2a II, and 16-cc 2ab) markers. We created a putative GC surface molecule
615 list and an EC surface molecule list. Next, we scanned the database generated by Guruharsha et
616 al. 2011 to identify which entries from the GC surface molecule list are predicted to form
617 physical protein:protein interaction pairs with entries from the EC surface molecule list. Next,
618 we explored GO terms associated with genes forming pairs. In Supplemental Table 4, we
619 shaded in yellow entries with GO term 'extracellular space', in blue with GO term 'plasma
620 membrane', and green if both GO terms are associated with the gene.

621 ## Monocle pseudotime analyses

622 We used Monocle 2 for pseudotime analyses of FSC differentiation (Qiu et al. 2017) using
623 standard workflow.

624 ## Dataset comparison

625 To compare the five datasets, we used Seurat v4 (Hao et al. 2020) data integration with
626 SCTransform. For marker comparison, we first computed positive markers using Wilcox test on
627 all datasets individually (Rust et al, Jevitt et al, Shi et al, Tu et al, and ours). For our datasets, we
628 merged all data into a single dataset. We omitted markers with adjusted p-value higher than
629 0.01, and selected 50 markers with lowest p-values for each cluster. Note: some clusters had
630 fewer than 50 markers. Next, we performed a pairwise comparison of marker overlap between
631 the clusters, and using Fisher's exact test to calculate corrected p-values. On the dot plot we
632 displayed % of marker overlap for each cluster, and $-\log_{10}()$ of corrected p-value. For detailed
633 discussion of data set comparison see Figures S8-10 and Supplemental Text.

634 ## RNA *in situ* hybridization and immunofluorescence.

635 HCR protocol was adapted from Choi *et al*. (Choi et al. 2018). Custom designed probes, probe
636 hybridization buffer, probe wash buffer and amplification buffer were procured from Molecular
637 Instruments, Inc.

638 # Data access

639 All raw and processed sequencing data generated in this study have been submitted to the
640 NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession
641 number GSE162192.

642 # Competing interest statement

643 Authors declare no competing interests.

644 ## Acknowledgements

645 The alpha-spectrin antibody developed by D. Branton and R. Dubreuil was obtained from the
646 Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at
647 the University of Iowa, Department of Biology, Iowa City, IA 52242. pMad antibody from Dr.
648 Edward Laufer. We thank Drs. Katja Rust and Amy Reilein for help identifying follicle stem cells
649 in stained germaria. We thank Drs. Katja Rust and Todd Nystul for sharing their results prior to
650 publication. We thank Drs. Katja Rust and Todd Nystul, Deeptiman Chatterjee and Wu-Min
651 Deng, Shiyuan Chen and Ting Xie, and Rongwen Xi for sharing annotated datasets. We thank
652 Drs. Sherilyn Grill and Lionel Christiaen for critical comments on the manuscript. We would like
653 to thank the Genome Technology Center (GTC) for expert library preparation and sequencing.
654 GTC is a shared resource partially supported by the Cancer Center Support Grant P30CA016087
655 at the Laura and Isaac Perlmutter Cancer Center. Cell sorting technologies were provided by
656 NYU Langone's Cytometry and Cell Sorting Laboratory, which is supported in part by grant
657 P30CA016087 from the National Institutes of Health/National Cancer Institute. SG was
658 supported by Dean's Undergraduate Research Fund Grant, RL was supported by the Simons
659 Foundation, NIH R37HD41900 and was an HHMI investigator. Author contributions: MS and RL
660 conceived and designed the study; MS, SG and TUB performed experiments; MS collected and
661 analyzed the data, and MS and RL wrote and the manuscript.

662

663 ## References

664 Allbee AW, Rincon-Limas DE, Biteau B. 2018. Lmx1a is required for the development of the
665 ovarian stem cell niche in Drosophila. *Development* **145**: dev163394.

666 Assa-Kunik, E., Torres, I.L., Schejter, E.D., Johnston, D.S., and Shilo, B.-Z. 2007. Drosophila follicle
667 cells are patterned by multiple levels of Notch signaling and antagonism between the Notch
668 and JAK/STAT pathways. *Development* **134**, 1161–1169.

669 Bai J, Montell D. 2002. Eyes absent, a key repressor of polar cell fate during Drosophila
670 oogenesis. *Development* **129**: 5377–5388.

671 Banisch TU, Maimon I, Dadosh T, Gilboa L. 2017. Escort cells generate a dynamic compartment
672 for germline stem cell differentiation via combined Stat and Erk signalling. *Development*
673 **144**: 1937–1947.

674 Beccari S, Teixeira L, Rørth P. 2002. The JAK/STAT pathway is required for border cell migration
675 during Drosophila oogenesis. *Mechanisms of Development* **111**: 115–123.

676 Bickel D, Shah R, Gesualdi SC, Haerry TE. 2008. Drosophila Follistatin exhibits unique structural
677 modifications and interacts with several TGF-beta family members. *Mechanisms of*
678 *Development* **125**: 117–129.

679 Bolívar J, Pearson J, López-Onieva L, González-Reyes A. 2006. Genetic dissection of a stem cell
680 niche: the case of the Drosophila ovary. *Developmental Dynamics* **235**: 2969–2979.

681 Boyle, M., and DiNardo, S. 1995. Specification, migration and assembly of the somatic cells of
682 the Drosophila gonad. *Development* **121**, 1815–1825.

683 Brigaud I, Duteyrat J-L, Chlasta J, Bail SL, Couderc J-L, Grammont M. 2015. Transforming Growth
684 Factor β/activin signalling induces epithelial cell flattening during Drosophila oogenesis.
685 *Biology open* **4**: 345–354.

686 Chang Y-C, Jang ACC, Lin C-H, Montell DJ. 2013. Castor is required for Hedgehog-dependent
687 cell-fate specification and follicle stem cell maintenance in Drosophila oogenesis.
688 *Proceedings of the National Academy of Sciences of the United States of America* **110**:
689 E1734-42.

690 Chen, D., and McKearin, D.M. 2003. A discrete transcriptional silencer in the *bam* gene
691 determines asymmetric division of the Drosophila germline stem cell. *Development* **130**,
692 1159–1170.

693 Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, Pierce
694 NA. 2018. Third-generation in situ hybridization chain reaction: multiplexed, quantitative,
695 sensitive, versatile, robust. *Development* **145**: dev165753.

696 Clarkson M, Saint R. 1999. A His2AvDGFP fusion gene complements a lethal His2AvD mutant
697 allele and provides an in vivo marker for Drosophila chromosome behavior. *DNA and cell*
698 *biology* **18**: 457–462.

699 Cooley L, Verheyen E, Ayers K. 1992. chickadee encodes a profilin required for intercellular
700 cytoplasm transport during Drosophila oogenesis. *Cell* **69**: 173–184.

701 Decotto E, Spradling AC. 2005. The Drosophila ovarian and testis stem cell niches: similar
702 somatic stem cells and signals. *Developmental Cell* **9**: 501–510.

703 Deng WM, Althauser C, Ruohola-Baker H. 2001. Notch-Delta signaling induces a transition from
704 mitotic cell cycle to endocycle in Drosophila follicle cells. *Development* **128**: 4737–4746.

705 Drummond-Barbosa D, Spradling AC. 2001. Stem Cells and Their Progeny Respond to
706 Nutritional Changes during Drosophila Oogenesis. *Developmental Biology* **231**: 265–278.
707 http://www.ncbi.nlm.nih.gov/pubmed/11180967.

708 Duhart JC, Parsons TT, Raftery LA. 2017. The repertoire of epithelial morphogenesis on display:
709 Progressive elaboration of Drosophila egg structure. *Mechanisms of Development* **148**: 18–
710 39.

711 Eliazer, S., Palacios, V., Wang, Z., Kollipara, R.K., Kittler, R., and Buszczak, M. 2014. Lsd1 Restricts
712 the Number of Germline Stem Cells by Regulating Multiple Targets in Escort Cells. *PLoS*
713 *Genet* **10**, e1004200.

714 Fadiga J, Nystul TG. 2019. The follicle epithelium in the Drosophila ovary is maintained by a
715 small number of stem cells. *eLife* **8**: 1161.

716 Flaherty MS, Salis P, Evans CJ, Ekas LA, Marouf A, Zavadil J, Banerjee U, Bach EA. 2010. chinmo
717 is a functional effector of the JAK/STAT pathway that regulates eye development, tumor
718 formation, and stem cell self-renewal in Drosophila. *Developmental Cell* **18**: 556–568.

719 Forbes AJ, Spradling AC, Ingham PW, Lin H. 1996. The role of segment polarity genes during
720 early oogenesis in Drosophila. *Development* **122**: 3283–3294.

721 Fuller, M.T., and Spradling, A.C. 2007. Male and Female Drosophila Germline Stem Cells: Two
722 Versions of Immortality. *Science* **316**, 402–404.

723 Godt, D., and Laski, F.A. 1995. Mechanisms of cell rearrangement and cell recruitment in
724 Drosophila ovary morphogenesis and the requirement of bric à brac. *Development* **121**,
725 173–187.

726 González-Reyes A, Elliott H, Johnston DS. 1995. Polarization of both major body axes in
727 Drosophila by gurken-torpedo signalling. *Nature* **375**: 654–658.

728 Guruharsha KG, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY,
729 Cenaj O, et al. 2011. A Protein Complex Network of Drosophila melanogaster. *Cell* **147**: 690–
730 703.

731 Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J.,
732 Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell
733 184, 3573–3587.e29.

734 Hartman TR, Ventresca EM, Hopkins A, Zinshteyn D, Singh T, O'Brien JA, Neubert BC, Hartman
735 MG, Schofield HK, Stavrides KP, et al. 2015. Novel tools for genetic manipulation of follicle
736 stem cells in the Drosophila ovary reveal an integrin-dependent transition from quiescence
737 to proliferation. *Genetics* **199**: 935–957.

738 Hill JH, Chen Z, Xu H. 2014. Selective propagation of functional mitochondrial DNA during
739 oogenesis restricts the transmission of a deleterious mitochondrial variant. *Nature Genetics*
740 **46**: 389–392.

741 Hu Y, Comjean A, Perkins LA, Perrimon N, Mohr SE. 2015. GLAD: an Online Database of Gene
742 List Annotation for Drosophila. *Journal of genomics* **3**: 75–81.

743 Huynh J-R. 2006. Fusome as a Cell-Cell Communication Channel of Drosophila Ovarian Cyst.
744 *Cell-Cell Channels*, pp. 217–235, Springer New York.

745 Jackson SM, Blochlinger K. 1997. cut interacts with Notch and protein kinase A to regulate egg
746 chamber formation and to maintain germline cyst integrity during Drosophila oogenesis.
747 *Development* **124**: 3663–3672.

748 Jevitt A, Chatterjee D, Xie G, Wang X-F, Otwell T, Huang Y-C, Deng W-M. 2020. A single-cell atlas
749 of adult Drosophila ovary identifies transcriptional programs and somatic cell lineage
750 regulating oogenesis. *PLoS Biology* **18**: e3000538.

751 Jordan KC, Clegg NJ, Blasi JA, Morimoto AM, Sen J, Stein D, McNeill H, Deng WM, Tworoger M,
752 Ruohola-Baker H. 2000. The homeobox gene mirror links EGF signalling to embryonic dorso-
753 ventral axis formation through notch activation. *Nature Genetics* **24**: 429–433.

754 Kai T, Williams D, Spradling AC. 2005. The expression profile of purified Drosophila germline
755 stem cells. *Developmental Biology* **283**: 486–502.

756 Kakugawa S, Langton PF, Zebisch M, Howell S, Chang T-H, Liu Y, Feizi T, Bineva G, O'Reilly N,
757 Snijders AP, et al. 2015. Notum deacylates Wnt proteins to suppress signalling activity.
758 *Nature* **519**: 187–192.

759 Kawashima T, Nakamura A, Yasuda K, Kageyama Y. 2003. Dmaf, a novel member of Maf
760 transcription factor family is expressed in somatic gonadal cells during embryonic
761 development and gametogenesis in Drosophila. *Gene Expression Patterns* **3**: 663–667.

762 King, R.C., (1970) Ovarian Development in *Drosophila melanogaster*, New York: *Academic*.

763 Kirilly, D., and Xie, T. 2007. The Drosophila ovary: an active stem cell community. **Cell Res. 17**,
764 15–25.

765 Kirilly D, Wang S, Xie T. 2011. Self-maintained escort cells form a germline stem cell
766 differentiation niche. *Development* **138**: 5087–5097.

767 Lasko PF, Ashburner M. 1988. The product of the Drosophila gene vasa is very similar to
768 eukaryotic initiation factor-4A. *Nature* **335**: 611–617.

769 Li MA, Alls JD, Avancini RM, Koo K, Godt D. 2003. The large Maf factor Traffic Jam controls
770 gonad morphogenesis in Drosophila. *Nature Cell Biology* **5**: 994–1000.

771 Lieber T, Jeedigunta SP, Palozzi JM, Lehmann R, Hurd TR. 2019. Mitochondrial fragmentation
772 drives selective removal of deleterious mtDNA in the germline. *Nature* **319**: 958.

773 Lin H, Spradling AC. 1995. Fusome asymmetry and oocyte determination in Drosophila.
774 *Developmental genetics* **16**: 6–12.

775 Liu Z, Zhong G, Chai PC, Luo L, Liu S, Yang Y, Baeg G-H, Cai Y. 2015. Coordinated niche-
776 associated signals promote germline homeostasis in the Drosophila ovary. *Journal of Cell*
777 *Biology* **211**: 469–484. http://jcb.rupress.org/content/211/2/469.

778 Lomas MF, Hails F, Lachance J-FB, Nilson LA. 2013. Response to the dorsal anterior gradient of
779 EGFR signaling in Drosophila oogenesis is prepatterned by earlier posterior EGFR activation.
780 *Cell reports* **4**: 791–802.

781 Lomas MF, Vito SD, Lachance J-FB, Houde J, Nilson LA. 2016. Determination of EGFR Signaling
782 Output by Opposing Gradients of BMP and JAK/STAT Activity. *Current biology: CB* **26**:
783 2572–2582.

784 Mahowald, A., Kambysellis. M. 1980. Oogenesis, M. Ashburner, T.R.F. Wright (Eds.), The
785 Genetics and Biology of *Drosophila*, Vol. 2d, Academic Press, New York (1980), pp. 141-224

786 Margaritis, L.H., Kafatos, F.C., and Petri, W.H. 1980. The eggshell of Drosophila melanogaster. I.
787 Fine structure of the layers and regions of the wild-type eggshell. *J Cell Sci* **43**, 1–35.

788 Margolis, J., and Spradling, A. 1995. Identification and behavior of epithelial stem cells in the
789 Drosophila ovary. *Development* **121**, 3797–3807.

790 Moore, L.A., Broihier, H.T., Van Doren, M., Lunsford, L.B., and Lehmann, R. 1998. Identification
791 of genes controlling germ cell migration and embryonic gonad formation in Drosophila.
792 *Development* **125**, 667–678.

793 Morimoto AM, Jordan KC, Tietze K, Britton JS, O'Neill EM, Ruohola-Baker H. 1996. Pointed, an
794 ETS domain transcription factor, negatively regulates the EGF receptor pathway in
795 Drosophila oogenesis. *Development* **122**: 3745–3754.

796 Mottier-Pavie VI, Palacios V, Eliazer S, Scoggin S, Buszczak M. 2016. The Wnt pathway limits
797 BMP signaling outside of the germline stem cell niche in Drosophila ovaries. *Developmental*
798 *Biology* **417**: 50–62.

799 Murphy AM, Lee T, Andrews CM, Shilo BZ, Montell DJ. 1995. The breathless FGF receptor
800 homolog, a downstream target of Drosophila C/EBP in the developmental control of cell
801 migration. *Development* **121**: 2255–2263.

802 Nelson N, Perzov N, Cohen A, Hagai K, Padler V, Nelson H. 2000. The cellular biology of proton-
803 motive force generation by V-ATPases. *The Journal of experimental biology* **203**: 89–95.

804 Nystul T, Spradling A. 2007. An epithelial niche in the Drosophila ovary undergoes long-range
805 stem cell replacement. *Cell Stem Cell* **1**: 277–285.

806 Nystul T, Spradling A. 2010. Regulation of epithelial stem cell replacement and follicle formation
807 in the Drosophila ovary. *Genetics* **184**: 503–515.

808 Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C. 2017. Reversed graph
809 embedding resolves complex single-cell trajectories. *Nature Methods* **14**: 979–982.

810 Reilein A, Melamed D, Park KS, Berg A, Cimetta E, Tandon N, Vunjak-Novakovic G, Finkelstein S,
811 Kalderon D. 2017. Alternative direct stem cell derivatives defined by stem cell location and
812 graded Wnt signalling. *Nature Cell Biology* **19**: 433–444.

813 Riechmann, V., Rehorn, K.P., Reuter, R., and Leptin, M. 1998. The genetic control of the
814 distinction between fat body and gonadal mesoderm in Drosophila. *Development* **125**, 713–
815 723.

816 Rojas-Rios, P., Guerrero, I., and Gonzalez-Reyes, A. 2012. Cytoneme-mediated delivery of
817 hedgehog regulates the expression of bone morphogenetic proteins to maintain germline
818 stem cells in Drosophila. *PLoS Biol* **10**, e1001298.

819 Rust, K., Byrnes, L.E., Yu, K.S., Park, J.S., Sneddon, J.B., Tward, A.D., and Nystul, T.G. (2020). A
820 single-cell atlas and lineage analysis of the adult Drosophila ovary. *Nat Comms 11*, 5628–17.

821 Sahai-Hernandez, P., and Nystul, T.G. (2013). A dynamic population of stromal cells contributes
822 to the follicle stem cell niche in the Drosophila ovary. Development *140*, 4490–4498.

823 Sahut-Barnola, I., Godt, D., Laski, F.A., and Couderc, J.L. 1995. Drosophila ovary morphogenesis:
824 analysis of terminal filament formation and identification of a gene required for this
825 process. *Developmental Biology* **170**, 127–135.

826 Santos GD, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB,
827 Gelbart WM, Consortium F. 2015. FlyBase: introduction of the Drosophila melanogaster
828 Release 6 reference genome assembly and large-scale migration of genome annotations.
829 *Nucleic Acids Research* **43**: D690-7.

830 Schulz C, Wood CG, Jones DL, Tazuke SI, Fuller MT. 2002. Signaling from germ cells mediated by
831 the rhomboid homolog stet organizes encapsulation by somatic support cells. *Development*
832 **129**: 4523–4534.

833 Shi, J., Jin, Z., Yu, Y., Zhang, Y., Yang, F., Huang, H., Cai, T., and Xi, R. 2020. A Progressive Somatic
834 Cell Niche Regulates Germline Cyst Differentiation in the Drosophila Ovary. *Curr. Biol.* **31**,
835 840–852.e845.

836 Silver DL, Montell DJ. 2001. Paracrine signaling through the JAK/STAT pathway activates
837 invasive behavior of ovarian epithelial cells in Drosophila. *Cell* **107**: 831–841.

838 Slaidina M, Banisch TU, Gupta S, Lehmann R. 2020. A single-cell atlas of the developing
839 Drosophila ovary identifies follicle stem cell progenitors. *Genes & Development* **34**: 239–
840 249.

841 Slaidina M, Lehmann R. 2014. Translational control in germline stem cell development. *The*
842 *Journal of Cell Biology* **207**: 13–21.

843 Song, X., Wong, M.D., Kawase, E., Xi, R., Ding, B.C., McCarthy, J.J., and Xie, T. 2004. Bmp signals
844 from niche cells directly repress transcription of a differentiation-promoting gene, bag of
845 marbles, in germline stem cells in the Drosophila ovary. *Development* **131**, 1353–1364.

846 Song X, Xie T. 2003. Wingless signaling regulates the maintenance of ovarian somatic stem cells
847 in Drosophila. *Development* **130**: 3259–3268.

848 Song, X., Zhu, C.-H., Doan, C., and Xie, T. 2002. Germline stem cells anchored by adherens
849 junctions in the Drosophila ovary niches. *Science* **296**, 1855–1857.

850 Spradling AC. 1993. Developmental genetics of oogenesis. *Developmental Genetics of*
851 *Oogenesis* 1–70.

852 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M,
853 Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**: 1888-
854 1902.e21.

855 Sun J, Deng W-M. 2007. Hindsight mediates the role of notch in suppressing hedgehog signaling
856 and cell proliferation. *Developmental Cell* **12**: 431–442.

857 Sun J, Deng W-M. 2005. Notch-dependent downregulation of the homeodomain gene cut is
858 required for the mitotic cycle/endocycle switch and cell differentiation in Drosophila follicle
859 cells. *Development* **132**: 4299–4308.

860 Timmons, A.K., Mondragon, A.A., Schenkel, C.E., Yalonetskaya, A., Taylor, J.D., Moynihan, K.E.,
861 Etchegaray, J.I., Meehan, T.L., and McCall, K. (2016). Phagocytosis genes nonautonomously
862 promote developmental cell death in the Drosophila ovary. *Proc. Natl. Acad*. Sci. U.S.a. **113**,
863 E1246–E1255.

864 Tootle TL, Williams D, Hubb A, Frederick R, Spradling A. 2011. Drosophila eggshell production:
865 identification of new genes and coordination by Pxt. ed. E.S.F.W. Abdelhay. *PloS one* **6**:
866 e19943.

867 Tu, R., Duan, B., Song, X., Chen, S., Scott, A., Hall, K., Blanck, J., DeGraffenreid, D., Li, H., Perera,
868 A., et al. 2020. Multiple Niche Compartments Orchestrate Stepwise Germline Stem Cell
869 Progeny Differentiation. *Curr. Biol*. **31**, 827–839.e3.

870 Tworoger M, Larkin MK, Bryant Z, Ruohola-Baker H. 1999. Mosaic analysis in the drosophila
871 ovary reveals a common hedgehog-inducible precursor stage for stalk and polar cells.
872 *Genetics* **151**: 739–748.

873 Vied, C., Reilein, A., Field, N.S., and Kalderon, D. 2012. Regulation of Stem Cells by Intersecting
874 Gradients of Long-Range Niche Signals. *Developmental Cell* **23**, 836–848.

875 Wang X, Page-McCaw A. 2018. Wnt6 maintains anterior escort cells as an integral component
876 of the germline stem cell niche. *Development* **145**: dev158527.

877 Wu X, Tanwar PS, Raftery LA. 2008. Drosophila follicle cells: morphogenesis in an eggshell.
878 *Seminars in cell & developmental biology* **19**: 271–282.

879 Xi R, McGregor JR, Harrison DA. 2003. A gradient of JAK pathway activity patterns the anterior-
880 posterior axis of the follicular epithelium. *Developmental Cell* **4**: 167–177.

881 Xie T, Spradling AC. 1998. decapentaplegic is essential for the maintenance and division of
882 germline stem cells in the Drosophila ovary. *Cell* **94**: 251–260.

883 Xie, T., and Spradling, A.C. 2000. A niche maintaining germ line stem cells in the Drosophila
884 ovary. *Science* **290**, 328–330.

885 Zhang, Y., and Kalderon, D. 2001. Hedgehog acts as a somatic stem cell factor in the Drosophila
886 ovary. *Nature* **410**, 599–604.

887 ## Figure Legends

888 **Figure 1. scRNA-seq of *Drosophila* adult ovaries.**

889 A - Schematic of the scRNA-seq sample preparation. B - UMAP plot of the entire dataset. Each
890 dot represents a transcriptome of a single cell, and is color-coded according to cluster
891 membership. C - Schematic drawing of an adult ovary. D - A heatmap visualizing average gene
892 expression levels of all marker genes in each cluster. Red indicates highest and blue lowest
893 expression. E - Expression of major cell type markers *vas, en, Wnt4, tj, Fas3, eya, cas* and *upd1*.
894 Red indicates highest, and light grey lowest expression. Arrowhead indicates *en* expressing
895 cells.

896 **Figure 2. GC subclustering in 9 differentiation states.**

897 A - UMAP plot of GCs subclustered in 9 differentiation states. Each dot represents a
898 transcriptome of a single cell, and is color-coded according to cluster membership. B -
899 Schematic drawing of a germarium and stage 2 follicle with GCs highlighted in the colors
900 corresponding to the UMAP plot in A. Germ cell differentiation stages (GSC, CB, and cyst stage)
901 are indicated in each cell. TFs, CCs, ECs and FC lineages are colored in white. Germarium regions
902 are indicated above the cartoon. Color-coded boxes indicate select marker genes for each
903 cluster, whose mRNAs are visualized in 3D-J, M, N. Note: Contrary to the schematic drawing, *in*
904 *vivo* germ cell differentiation stages are not represented equally frequently. C - A dot plot
905 visualizing expression of 14 marker genes in 9 GC clusters. Dot diameter represents the fraction
906 of cells expressing each gene in each cluster, as shown in scale. Color intensity represents the
907 average normalized expression level. D-J, M, N - Marker gene mRNA *in situ* hybridization using
908 HCR (cyan, magenta). Spectrosomes, fusomes and somatic cell membranes are labeled by anti-
909 α-Spectrin (yellow). Scale bars - 10 μm. D - *CG15398* (cyan) and *CG2887* (magenta) are
910 expressed in GSCs, CBs and 2-cell cysts. E - *bam* (cyan) is predominantly expressed in 4-cell
911 cysts, *CG6628* (magenta) is expressed in GSCs, CBs and 2-cell cysts. F - *HP6* (cyan) is expressed
912 in GSCs, CBs, 2-cell and 4-cell cysts, *Orc1* (magenta) is expressed in 4-cell and 8-cell cysts. G -
913 *bam* (cyan) is predominantly expressed in 4-cell cysts, with lower expression in 2-cell and later
914 stage cysts, *CG15628* (magenta) is expressed in 8-cell and 16-cell cysts in germarium regions 2a
915 and 2b. H - *CG3961* (cyan) is expressed predominantly in 16-cell cysts in germarium regions 2a
916 and 2b. I - *Nlg2* (cyan) is expressed in 16-cell cysts in germarium regions 2a and 2b. J - *mnd*
917 (cyan) is expressed in 16-cell cysts in germarium regions 2a, 2b and 3, and in stage 2 follicles. K,
918 L - Heatmaps visualizing average gene expression levels of all GC marker genes (K) and
919 dynamically expressed transcription factors (L) in each cluster. Red indicates highest and blue
920 lowest expression. M - *CG17270* (cyan) is expressed in 16-cell cysts in germarium region 3 and
921 in stage 2 follicle, *CG7225* (magenta) is expressed in 16-cell cysts in germarium regions 2b and
922 3, in stage 2 follicle and beyond. N - *slif* (cyan) is expressed in stage 2 follicle and beyond, *cry*
923 (magenta) is expressed in 16-cell cysts in germarium regions 2b and 3, and in follicle cells.

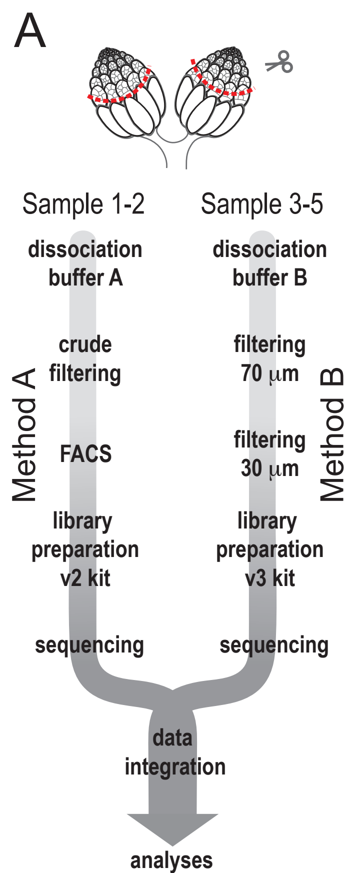
924 **Figure 3. Identification of three transcriptionally distinct EC subtypes.**

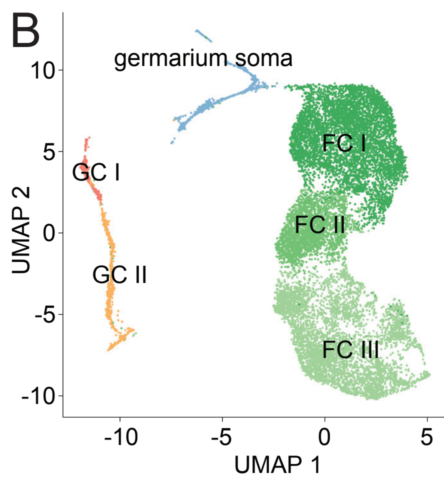
925 A - UMAP plot of somatic cells of the germarium subclustered in 8 clusters. Each dot represents
926 a transcriptome of a single cell, and is color-coded according to cluster membership. B -
927 Schematic drawing of a germarium and stage 2 follicle with somatic cells highlighted in the
928 colors corresponding to the UMAP plot in A. GC are colored in white, and their differentiation
929 stages (GSC, CB, and cyst stage) are indicated in each cell. Germarium regions are indicated
930 above the cartoon. Color-coded boxes indicate select marker genes for each cluster. Note:
931 Contrary to the schematic drawing, *in vivo* germ cell differentiation stages are not represented
932 equally frequently. C - A dot plot visualizing expression of select 28 marker genes. Dot diameter
933 represents the fraction of cells expressing each gene in each cluster, as shown in scale. Color
934 intensity represents the average normalized expression level. D-I, K-L, N-O - Marker gene mRNA
935 *in situ* hybridization using HCR (cyan, magenta, and yellow). Scale bars, 10 μm. g - indicates
936 germaria, TF - terminal filament, St2, St4, and St6 indicates oogenesis stages for follicles. E, G, H
937 - Spectrosomes, fusomes and somatic cell membranes are labeled by anti-α-Spectrin (yellow). D
938 - *Wnt4* (cyan) is expressed in TF, CC and ECs, *Lmx1a* (magenta) is expressed in TF and CC. E - *ptc*
939 (cyan) is expressed in ECs, *hh* (magenta) is expressed in ECs and pre-FCs. F - *NetA* (cyan) and
940 Notum (magenta) are expressed in anterior ECs in germarium region 1. G - *CG31431* (cyan) is
941 predominantly expressed in EC in germarium region 2a with some expression in region 1 and
942 2b, *wun2* (magenta) is predominantly expressed in EC in germarium region 2a. H - *kar* (cyan) is
943 sparsely and lowly expressed throughout the germarium, *CG10073* (magenta) is predominantly
944 expressed in region 2a ECs. I - *AdamTS-A* (cyan) is expressed in FSC, pre-FCs, and the most
945 posterior ECs, *wun2* (magenta) is predominantly expressed in EC in germarium region 2a, *laza*
946 (yellow) is expressed in anterior ECs in germarium region 1 and some ECs in germarium region
947 2a. J - A heatmap visualizing average expression levels of EC enriched secreted proteins and
948 ECM components. Red indicates highest and blue lowest expression. K - *AdamTS-A* (cyan) is
949 expressed in FSC, pre-FCs, and the most posterior ECs, *peb* (magenta) is sparsely and lowly
950 expressed throughout the germarium. L - *Fas3* (cyan) and *CG1136* (magenta) are expressed in
951 FC lineage including FSCs, pre-FCs, pre-stalk, and epithelial FCs. M - A heatmap visualizing
952 average gene expression levels of all somatic cell marker genes in each cluster. Red indicates
953 highest and blue lowest expression. N - *AdamTS-A* (cyan) is expressed in FC lineage including
954 FSCs, pre-FCs, pre-stalk, and epithelial FCs, *CG10311* (magenta) is expressed in stalk cells, *sim*
955 (yellow) expressed in pre-stalk and stalk cells. O - *CG34377* (cyan) and *Hk* (magenta) are
956 expressed in polar cells, Dapi (white) labels cell nuclei.

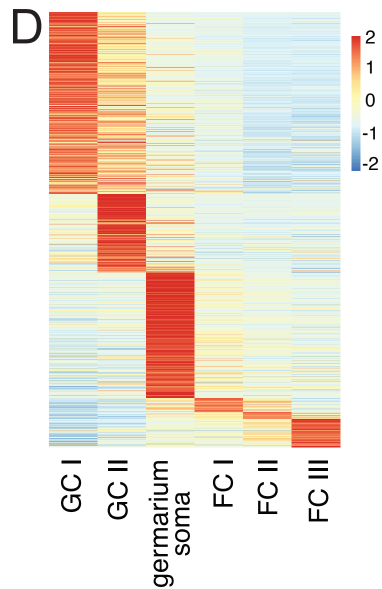
957 **Figure 4. Characterization of 17 FC subpopulations.**

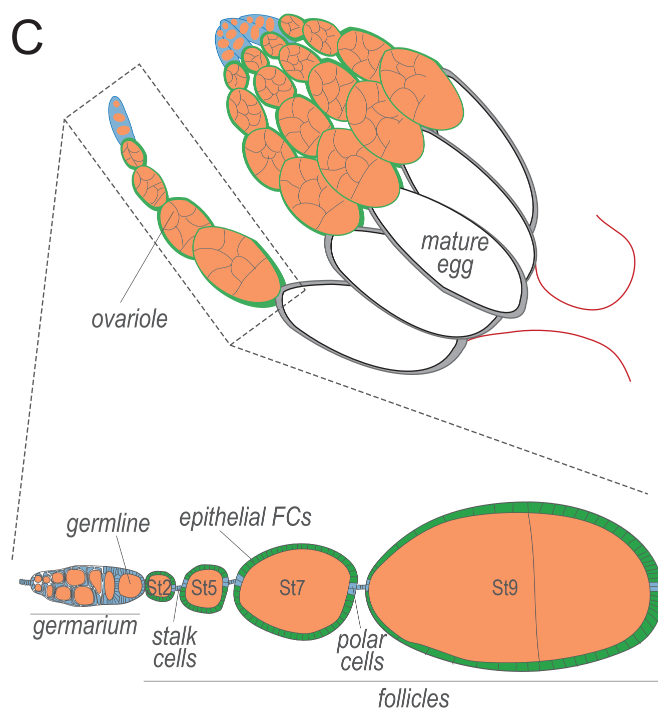
958 A - UMAP plot of FCs subclustered in 17 clusters. Each dot represents a transcriptome of a
959 single cell, and is color-coded according to cluster membership. B - A dot plot visualizing
960 expression of select 18 marker genes. Dot diameter represents the fraction of cells expressing
961 each gene in each cluster, as shown in scale. Color intensity represents the average normalized
962 expression level. C - A heatmap visualizing average gene expression levels of all FC marker
963 genes in each cluster. Red indicates highest and blue lowest expression. D - Schematic drawing
964 of an ovariole with FC subtypes highlighted in the colors corresponding to the UMAP plot in 4A.
965 Note: Contrary to the schematic drawing, *in vivo* follicles of all stages are not represented
966 equally frequently. E-N - Marker gene mRNA *in situ* hybridization using HCR (cyan, magenta,

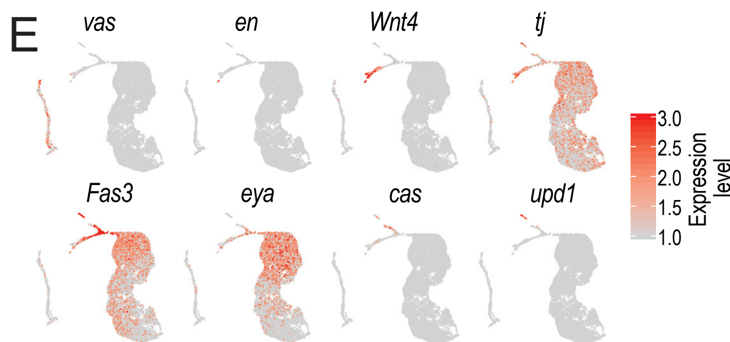
967 yellow). Follicle stages (2-9) and germarium (g) are indicated above or below each follicle. Scale
968 bars, 30 µm. E - *ImpL2* (cyan) is expressed in stretch cells at stage 9, *CG1136* (magenta) is
969 expressed in pre-FCs, and stage 2-4 FCs, *Asph* (yellow) expressed in stage 6 FC, and restricted to
970 both termini at stage 7. F - *CG31808* (cyan) is faintly expressed in stage 3-4 FCs, and is restricted
971 to ATFC at stage 6. G - *CG15546* (cyan) has patchy expression in stage 2-4 follicles, and is
972 restricted to PTFC by stage 5-6, *ct* (magenta) is expressed in all FCs from pre-FC to stage 6, *Asph*
973 (yellow) expressed in stage 5-6 follicle. H - *Spn42Dd* (cyan) is expressed in ATFCs at stage 6,
974 *CG10924* (magenta) is expressed in MBFCs at stage 6. I - *mirr* (cyan) is expressed in MBFC from
975 stage 6-7 onwards, *magu* (magenta) is expressed in ATFCs from stage 6 and is downregulated
976 by stage 9, *mid* (yellow) expressed in PTFC starting from stage 6. J - *ImpL2* (cyan) is expressed in
977 ATFC excluding the most anterior tip at stage 8 and in stretch and border cells at stage 9,
978 *CG5002* (magenta) is expressed in stretch cells at stage 9, *Amph* (yellow) expressed in ATFC at
979 sage 8. K - *mirr* (cyan) is expressed in MBFC from stage 6-7 onwards, *laza* (magenta) is weakly
980 expressed in PTFCs starting from stage 4, and strongly expressed in a broad band of PTFC at
981 stage 8, *H15* (yellow) is expressed in a narrow band of the most posterior PTFCs at stage 8,
982 weak expression is detectable in PTFCs at stage 6. L - *ImpL2* (cyan) is expressed in stretch cells
983 at stage 9, *bond* (magenta) is expressed in a posterior-to-anterior gradient in PTFCs and MBFCs
984 at stage 9. M - *CG8303* (magenta) is expressed in a posterior-to-anterior gradient in PTFCs and
985 MBFCs at stages 8 and 9. N - *ImpL2* (cyan) is expressed in ATFCs at stage 8, extending into
986 MBFCs, and in stretched cells at stage 9, *Yp2* (magenta) is expressed in a posterior-to-anterior
987 gradient in PTFCs and MBFCs at stages 8 and 9. Its expression is excluded from the most
988 posterior tip.
989


A
Sample 1-2
Sample 3-5
dissociation buffer A
dissociation buffer B
crude filtering
filtering 70 μm
FACS
filtering 30 μm
Method A
Method B
library preparation v2 kit
library preparation v3 kit
sequencing
sequencing
data integration
analyses

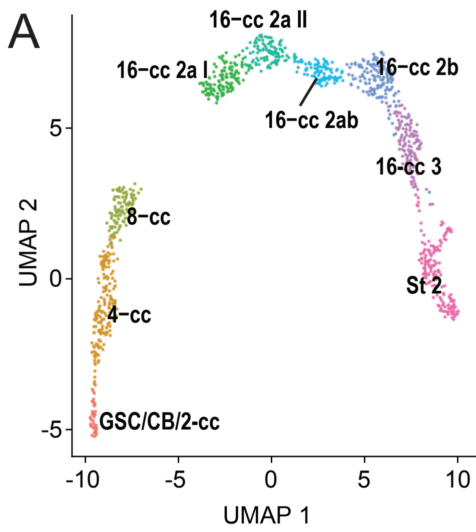

B
germarium soma
GC I
GC II
FC I
FC II
FC III
UMAP 2
UMAP 1
10
5
0
-5
-10
-10
-5
0
5

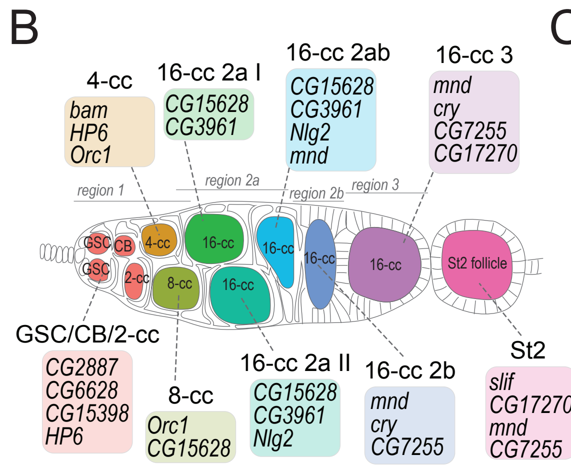

D
2
1
0
-1
-2
GC I
GC II
germarium soma
FC I
FC II
FC III

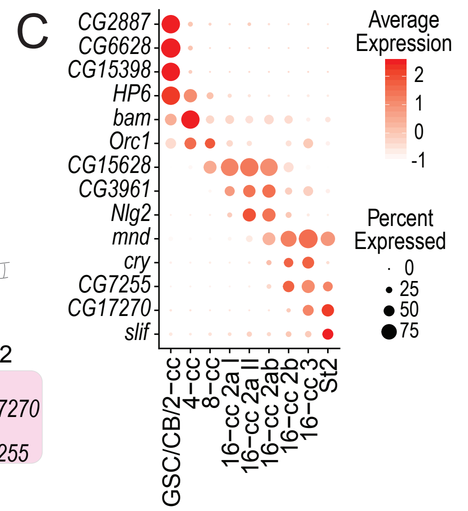

C
ovariole
mature egg

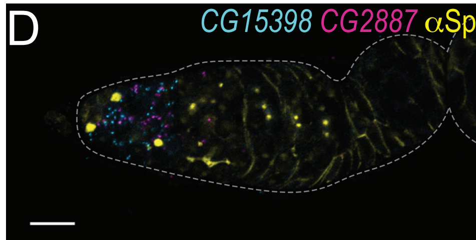

germline
epithelial FCs
St2
St5
St7
St9
germarium
stalk cells
polar cells
follicles

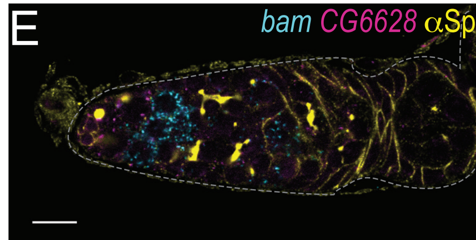
E
vas
en
Wnt4
tj
Fas3
eya
cas
upd1
3.0
2.5
2.0
1.5
1.0
Expression level

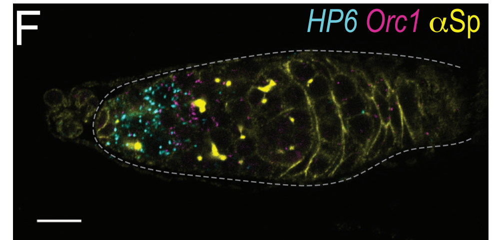

A
16−cc 2a II
16−cc 2a I
16−cc 2b
16−cc 2ab
16−cc 3
8−cc
St 2
4−cc
GSC/CB/2−cc
UMAP 2
UMAP 1

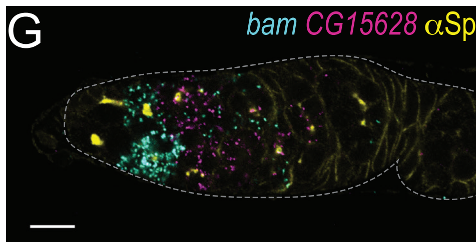

B
4-cc
bam
HP6
Orc1
16-cc 2a I
CG15628
CG3961
16-cc 2ab
CG15628
CG3961
Nlg2
mnd
16-cc 3
mnd
cry
CG7255
CG17270
region 1
region 2a
region 2b
region 3
St2 follicle
GSC/CB/2-cc
CG2887
CG6628
CG15398
HP6
8-cc
Orc1
CG15628
16-cc 2a II
CG15628
CG3961
Nlg2
16-cc 2b
mnd
cry
CG7255
St2
slif
CG17270
mnd
CG7255

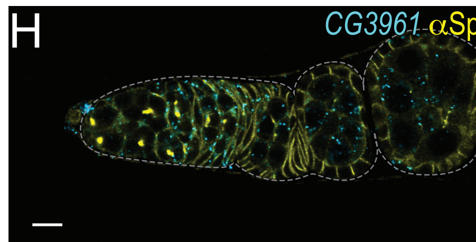

C
CG2887
CG6628
CG15398
HP6
bam
Orc1
CG15628
CG3961
Nlg2
mnd
cry
CG7255
CG17270
slif
Average Expression
Percent Expressed
GSC/CB/2−cc
4−cc
8−cc
16−cc 2a I
16−cc 2a II
16−cc 2ab
16−cc 2b
16−cc 3
St2

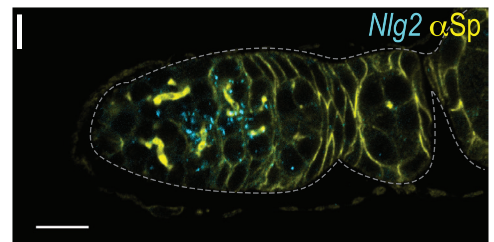

D
CG15398 CG2887 αSp

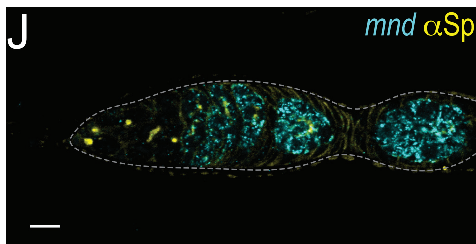

E
bam CG6628 αSp

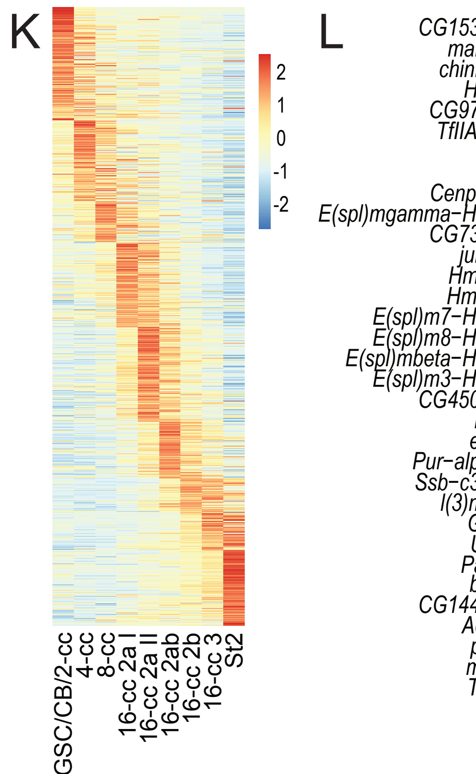

F
HP6 Orc1 αSp

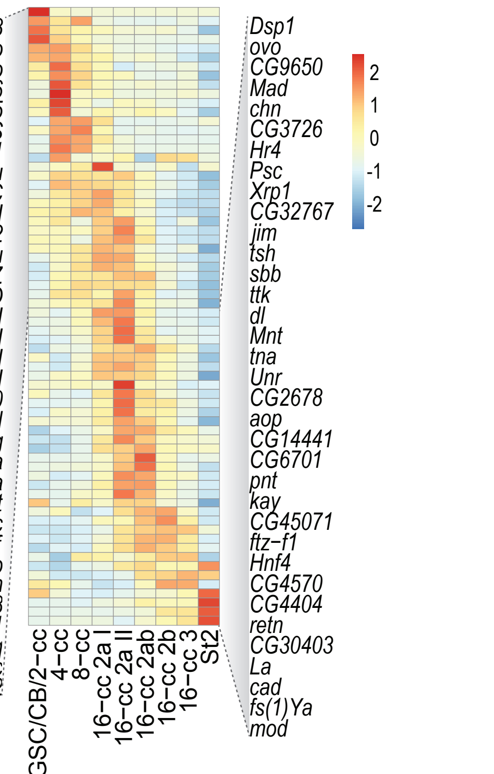

G
bam CG15628 αSp

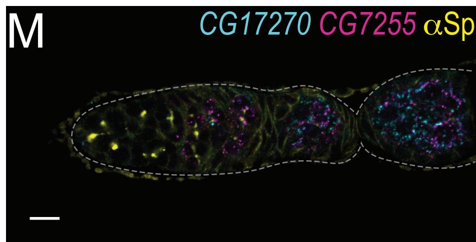

H
CG3961 αSp

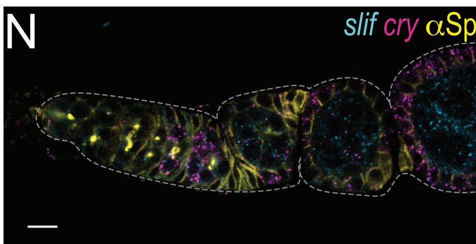

I
Nlg2 αSp

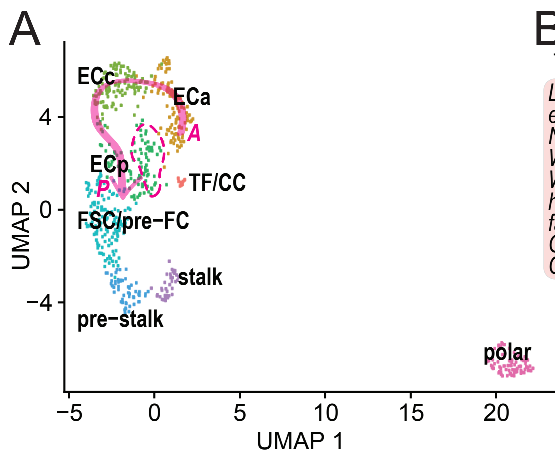
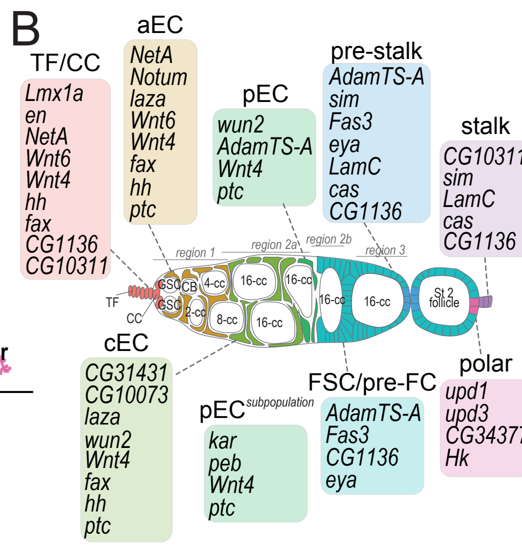
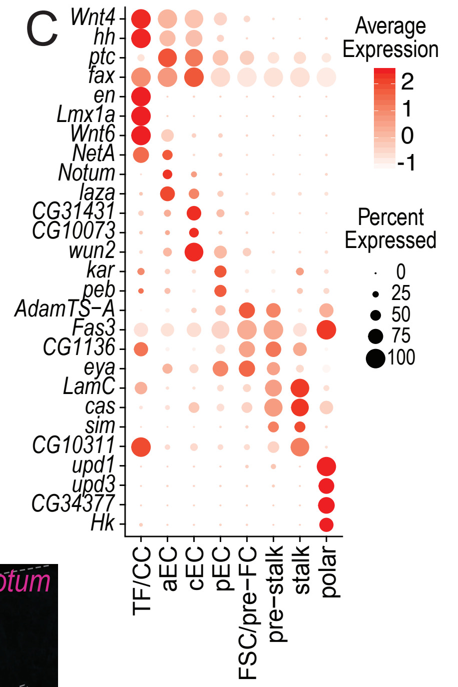
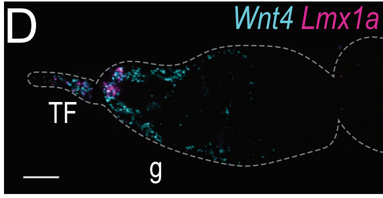
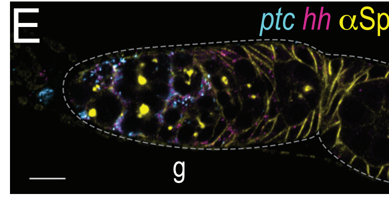
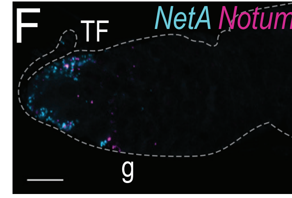
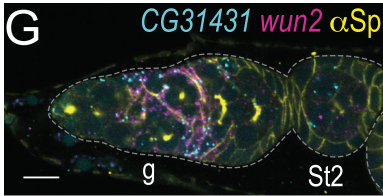
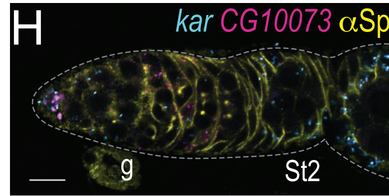
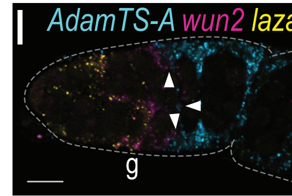
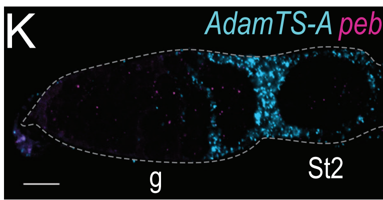
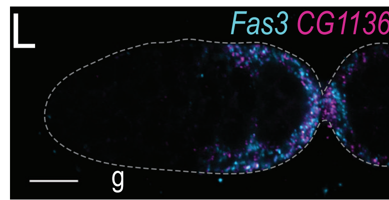
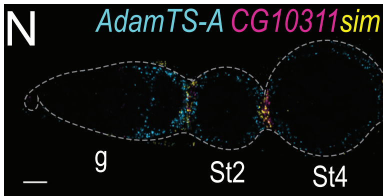
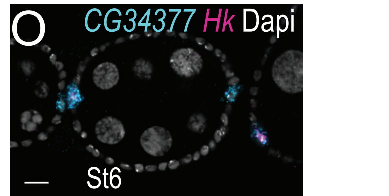
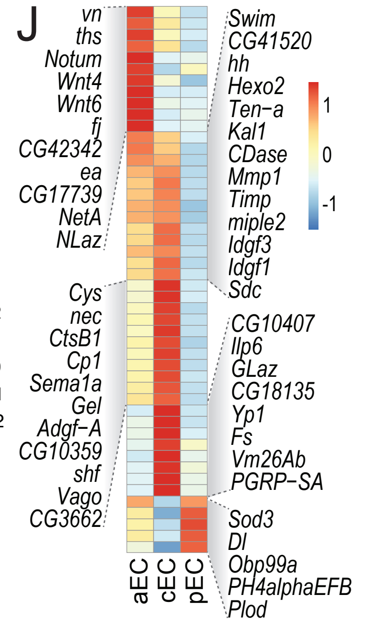
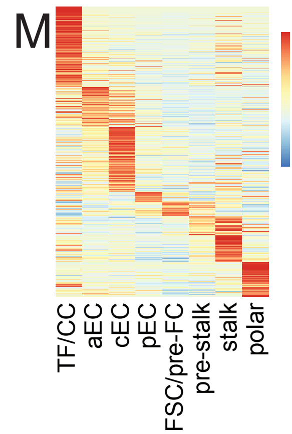

J
mnd αSp

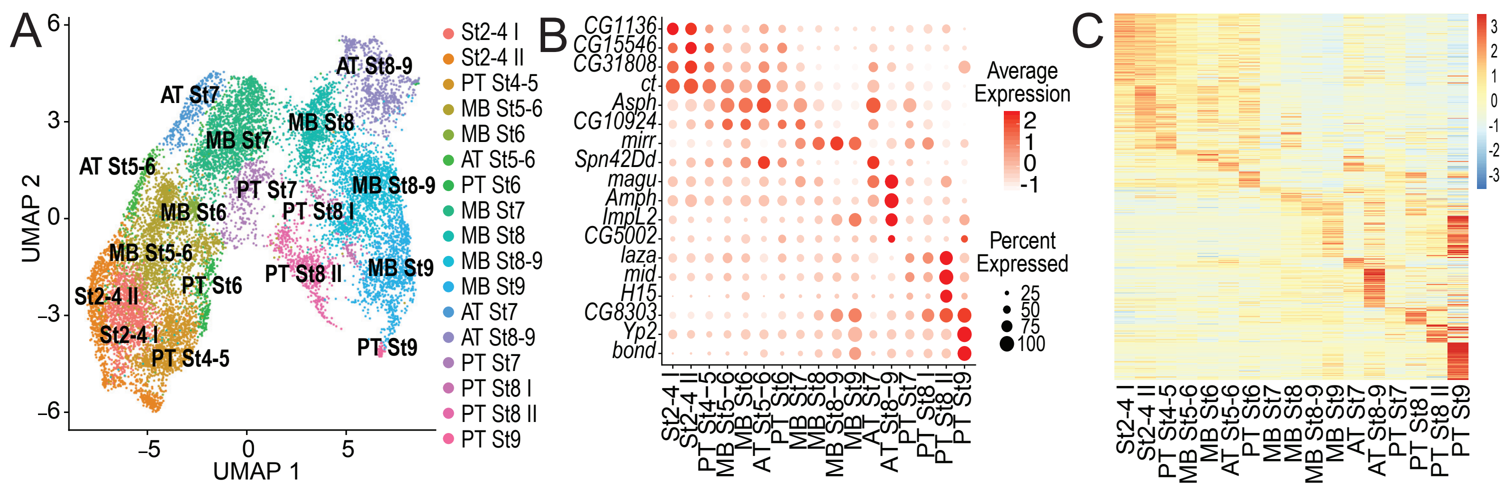

K
GSC/CB/2-cc
4-cc
8-cc
16-cc 2a I
16-cc 2a II
16-cc 2ab
16-cc 2b
16-cc 3
St2

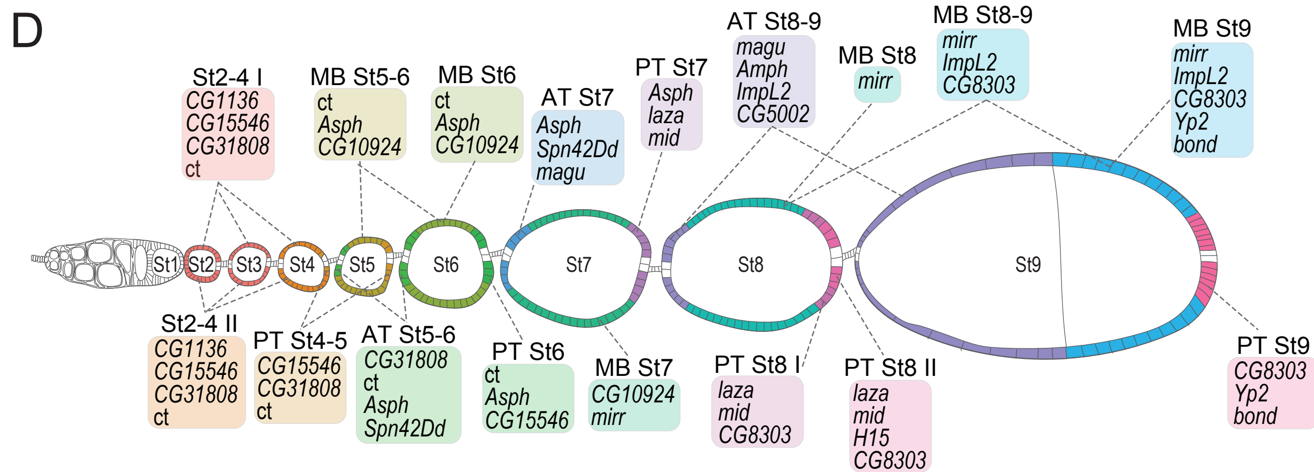

L
CG15398
mamo
chinmo
HP6
CG9705
TfIIA−S
D1
glu
Cenp−C
E(spl)mgamma−HLH
CG7372
jumu
HmgZ
HmgD
E(spl)m7−HLH
E(spl)m8−HLH
E(spl)mbeta−HLH
E(spl)m3−HLH
CG45050
Pcl
egg
Pur−alpha
Ssb−c31a
l(3)mbt
Gsc
Usf
Parp
bun
CG14438
Adf1
psq
mxc
Trf2
Dsp1
ovo
CG9650
Mad
chn
CG3726
Hr4
Psc
Xrp1
CG32767
jim
tsh
sbb
ttk
dl
Mnt
tna
Unr
CG2678
aop
CG14441
CG6701
pnt
kay
CG45071
ftz−f1
Hnf4
CG4570
CG4404
retn
CG30403
La
cad
fs(1)Ya
mod
GSC/CB/2−cc
4−cc
8−cc
16−cc 2a I
16−cc 2a II
16−cc 2ab
16−cc 2b
16−cc 3
St2

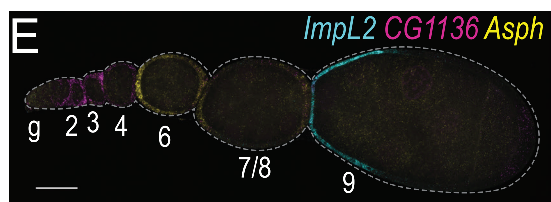

M
CG17270 CG7255 αSp

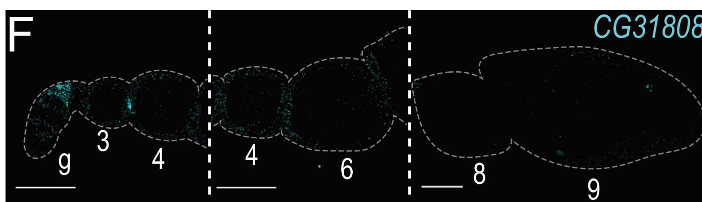

N
slif cry αSp

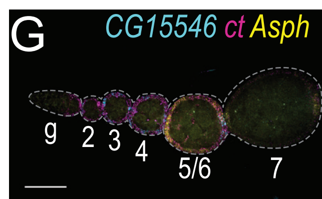

A
ECc
ECa
ECp
TF/CC
FSC/pre-FC
stalk
pre-stalk
polar
UMAP 2
UMAP 1

B
TF/CC
Lmx1a
en
NetA
Wnt6
Wnt4
hh
fax
CG1136
CG10311
aEC
NetA
Notum
laza
Wnt6
Wnt4
fax
hh
ptc
pEC
wun2
AdamTS-A
Wnt4
ptc
pre-stalk
AdamTS-A
sim
Fas3
eya
LamC
cas
CG1136
stalk
CG10311
sim
LamC
cas
CG1136
cEC
CG31431
CG10073
laza
wun2
Wnt4
fax
hh
ptc
pEC subpopulation
kar
peb
Wnt4
ptc
FSC/pre-FC
AdamTS-A
Fas3
CG1136
eya
polar
upd1
upd3
CG34377
Hk

C
Average Expression
Percent Expressed
TF/CC
aEC
cEC
pEC
FSC/pre-FC
pre-stalk
stalk
polar

D
Wnt4 Lmx1a
TF
g

E
ptc hh αSp
g

F
TF
NetA Notum
g

G
CG31431 wun2 αSp
g
St2

H
kar CG10073 αSp
g
St2

I
AdamTS-A wun2 laza
g

K
AdamTS-A peb
g
St2

L
Fas3 CG1136
g

N
AdamTS-A CG10311sim
g
St2
St4

O
CG34377 Hk Dapi
St6

J
aEC
cEC
pEC

M
TF/CC
aEC
cEC
pEC
FSC/pre-FC
pre-stalk
stalk
polar

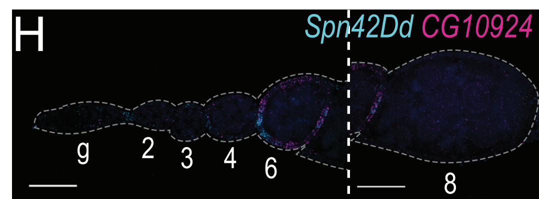

A
B
C
UMAP 1
UMAP 2
Average Expression
Percent Expressed
St2-4 I
St2-4 II
PT St4-5
MB St5-6
MB St6
AT St5-6
PT St6
MB St7
MB St8
MB St8-9
MB St9
AT St7
AT St8-9
PT St7
PT St8 I
PT St8 II
PT St9

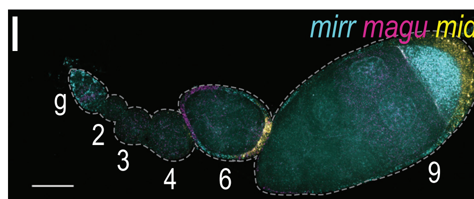

D
St2-4 I
CG1136
CG15546
CG31808
ct
MB St5-6
ct
Asph
CG10924
MB St6
ct
Asph
CG10924
AT St7
Asph
Spn42Dd
magu
PT St7
Asph
laza
mid
AT St8-9
magu
Amph
ImpL2
CG5002
MB St8
mirr
MB St8-9
mirr
ImpL2
CG8303
MB St9
mirr
ImpL2
CG8303
Yp2
bond
St2-4 II
CG1136
CG15546
CG31808
ct
PT St4-5
CG15546
CG31808
ct
AT St5-6
CG31808
ct
Asph
Spn42Dd
PT St6
ct
Asph
CG15546
MB St7
CG10924
mirr
PT St8 I
laza
mid
CG8303
PT St8 II
laza
mid
H15
CG8303
PT St9
CG8303
Yp2
bond

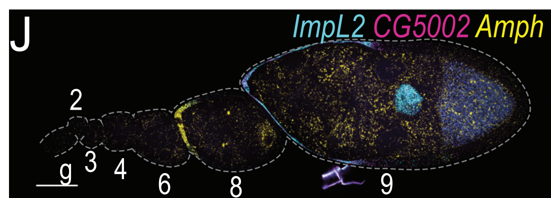

E
ImpL2 CG1136 Asph

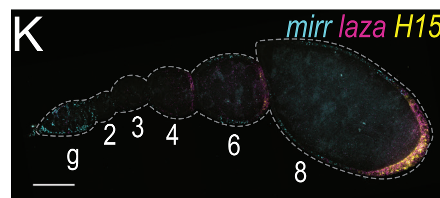

F
CG31808

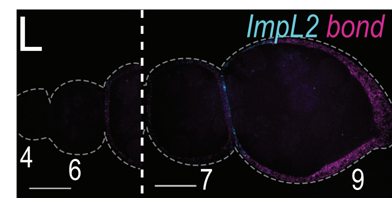

G
CG15546 ct Asph

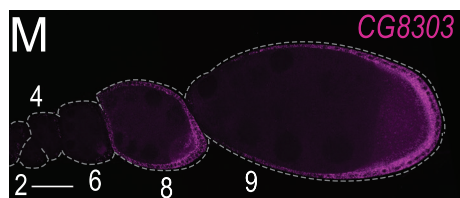

H
Spn42Dd CG10924

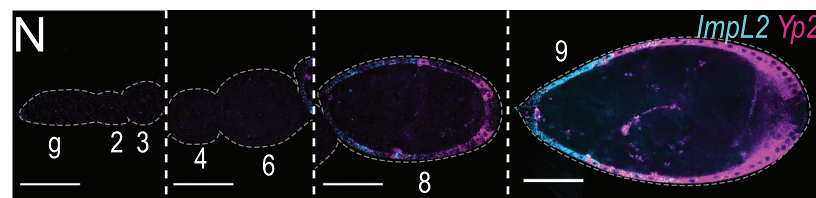

I
mirr magu mid


J
ImpL2 CG5002 Amph


K
mirr laza H15


L
ImpL2 bond


M
CG8303


N
ImpL2 Yp2