

## Determination of expression profiles for *Drosophila* ovarian Follicle Stem Cells (FSCs) using single-cell RNA sequencing

Zhi Dong<sup>1#</sup>, Lan Pang<sup>1#</sup>, Zhiguo Liu<sup>1</sup>, Yifeng Sheng<sup>1</sup>, Xavier Thibault<sup>2</sup>, Amy, Reilein<sup>2</sup>, Daniel Kalderon<sup>2\*</sup> and Jianhua Huang<sup>1\*</sup>

<sup>1</sup>Institute of Insect Sciences, Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insect Pests, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

<sup>2</sup>Department of Biological Sciences, Columbia University, New York, NY, USA

# These authors contributed equally

\* Corresponding authors: [ddk1@columbia.edu](mailto:ddk1@columbia.edu) (DK), [jhuang@zju.edu.cn](mailto:jhuang@zju.edu.cn) (JH)

### Summary

*Drosophila* ovarian Follicle Stem Cells (FSC) present an excellent paradigm for understanding how a community of active stem cells maintained by population asymmetry is regulated. Here we describe single-cell RNA sequencing studies of a pre-sorted population of cells that include FSCs and the neighboring cell types, Escort Cells (ECs) and Follicle Cells (FCs), which they support. Cell-type assignment relies on anterior-posterior (AP) location within the germarium. We clarify the previously determined location of FSCs and use spatially targeted lineage studies as further confirmation. The scRNA profiles among four clusters are consistent with an AP progression from anterior ECs through posterior ECs and then FSCs, to early FCs. Several genes with graded profiles from ECs to FCs are highlighted as candidate effectors of the inverse gradients of the two principal signaling pathways, Wnt and JAK-STAT, that guide FSC differentiation and division.

## Introduction

Adult stem cells provide a lifelong source of specific differentiated cells, necessitated by physiological turnover or changing environmental conditions (1-3). The study of adult stem cells is revealing that this task can be accomplished in a variety of ways. Importantly, it is often a shared task. In such paradigms, often termed population asymmetry, individual stem cells make stochastic decisions about cell division and differentiation (2-8). Division and differentiation are independent processes in the three best-studied paradigms (9-11) and are balanced at the community level. The stochastic variation in behavior of individual stem cells may be compounded further by systematic heterogeneity based on precise stem cell location (2, 12). For example, within a population of about 16 Follicle Stem Cells (FSCs) in a *Drosophila* germarium, roughly half are in a posterior ring. Only posterior FSCs can directly become proliferative Follicle Cells (FCs) and they also divide much faster than their anterior FSC neighbors, which can directly become quiescent Escort Cells (ECs) (13). Posterior and anterior FSCs nevertheless form a single community because they can exchange positions (13). It has also often been observed that stem cell derivatives can revert to stem cell status under physiological or stress conditions (14-16). All of these plastic properties, including variable stem cell lifetimes, differ substantially from the original concept, still relevant to some paradigms, of each stem cell behaving the same way- dividing with irreversible asymmetric outcomes for the two daughters and exhibiting exceptional longevity (1-3, 7).

Many aspects of the status of a cell can be captured by a detailed rendition of its pattern of RNA expression, so understanding of adult stem cell biology surely benefits from such information, typically captured by single-cell RNA (scRNA) sequencing. However, adult stem cell paradigms that exhibit stochastic and heterogeneous behaviors, with derivative cells that are necessarily initially very similar and do not necessarily quickly transit to an irreversibly distinct state, present a significant challenge to categorization of stem cells and their immediate derivatives according to scRNA profiles. These profiles cannot be expected to reveal stem cells as a highly distinct group and certainly cannot suffice to define stem cells. Instead, they must be related carefully to the results of functional tests that define stem cells through their behavior (2). *Drosophila* FSCs present a particularly attractive paradigm for such analyses because their

behavior is complex but has been studied carefully, along with extensive investigation of key regulatory niche signals (2, 9, 13, 17, 18). Here, we present a scRNA study of somatic ECs, FSCs and early FCs within the *Drosophila* germarium and we relate those studies to functional tests of FSC behavior based on location. The resulting expression profiles reveal candidate positionally-regulated effectors of FSC division and differentiation.

## Results

At the anterior of each germarium are non-dividing somatic terminal filament (TF) and cap cells, which contact 2-3 germline stem cells (GSCs) (Fig. 1A). GSCs generally divide asymmetrically to produce a more posterior Cystoblast, which divides four times with incomplete cytokinesis (19-21). The developing germline cyst derivatives are wrapped by processes of somatic ECs as they progress posteriorly through region 1 (22, 23) to form a rounded 16-cell stage 2a cyst, which then elongates into a lens-shaped 2b cyst, spanning the widest part of the germarium. Proliferative Follicle Cell precursors (“FCs”) associate with the 2b cyst, which then rounds to become a stage 3 cyst before budding from the germarium as an egg chamber, enclosed in a monolayer epithelium of FCs. Quiescent polar cells at the anterior and posterior termini of the FC epithelium contact non-dividing stalk cells, which connect consecutive egg chambers. Polar and stalk cell precursors are specified early within the germarium (24-26), while other FCs continue to divide until mid-oogenesis (stage 6), later specializing according to their AP and DV locations on the egg chamber (27, 28). FSCs lie between ECs and the earliest FCs, providing a continuous supply of new FCs and also, less frequently, replenishing ECs (13, 25).

Some scRNA studies have been conducted on whole ovarioles to capture all aspects of oogenesis (18, 29, 30). We were interested only in examining FSCs and their neighbors in detail, so we chose a strategy that began with purifying those cells. The enhancer trap line *C587-GAL4*, which is expressed in adult ECs, FSCs and the earliest FCs, was used to drive *UAS-CD8-RFP* cell surface protein expression, followed by manual dissection of ovaries, dissociation and fluorescence-activated cell sorting (FACS). Single-cell RNA sequences were derived on the 10XGenomics platform with unique molecular identifiers (UMIs). Average raw sequence reads

per cell were over 290,000. After genome alignment and UMI counting, genes featured in more than three cells were included, and cells with feature counts between 100 - 4,500, and fewer than 20% mitochondrial counts were retained for principal component analysis using Seurat (v4.0.2).

Following initial clustering of scRNA patterns for a total of over 1300 cells, we examined each group for known indicators of EC, FSC or early FC identity. We recognized small groups with features of cap cells or TF cells (selective expression of *Lmx1*, *en*, *dpp* and *wg*; 43 cells) (31-33), stalk cells and their precursors (selective expression of *LamC* and *sim*; 23 cells) (30), and germline cells (selective expression of *zpg*, *chinmo*, *ovo*, *otu* and *vasa*; 34 cells) (Fig. S1) (30). Three groups (0, 4, 9 in Fig. 2A) were not considered further on the basis of a combination of higher mitochondrial RNA content and lower total RNA counts, suggesting the possibility of damage or stress (Fig. 2B). The remaining four groups had characteristics of ECs and FSCs (*Wnt4*, *ptc*, *fax*) (13, 34-36) or early FCs (*Fas3*, *cas*) (24) and were re-sorted by themselves in order to gain better resolution among somatic cells of the anterior half of the germarium. The result was a progression of six clusters, roughly consistent with progressive anterior to posterior identities (Fig. 2C, D), as described further below.

We compared our clusters to those derived from another study that also concentrated on anterior somatic germarial cells (37) and found substantial similarities. The patterns of prominent selectively expressed RNAs that defined groups previously labeled as “IGS1-5” (IGS, or inner germarial sheath, is an alternative name for Escort Cells) showed largely good correspondence to our group 0 (IGS1 and 2; *NetA* + *Croc*), 5 (IGS3; *Croc* + *Hf*), 1 (IGS4; *bnb* + *wun2*) and 4 (IGS5; *wun2* + *santa-maria*) (Fig. 2E-G). Importantly, Tu et al., (37) used several of these “marker RNAs” for *in situ* hybridization to define the spatial limits of specific clusters, confirming the anterior to posterior progression of groups 0-1-5-4 in that order, as described below. Clusters 2 and 3 appear to represent more mature FCs, with ribosomal protein RNAs dominating the most up-regulated genes of cluster 2 (Fig. 3E), consistent with the highly proliferative nature of FCs. This feature is presaged in the earliest FCs in group 4 by strong expression of two ribosomal protein RNAs (*RpL35* and *RpS6*) and *myc* (Fig. 3D). A handful of

cells within group 2 expressed JAK-STAT ligands, *upd1* and *upd3*, characteristic of polar cells (38).

### **Cluster boundaries and cell identities**

A landmark that has been used to define FSCs by functional assays is the anterior border of strong surface Fas3 staining (Fig. 1A). In key functional studies this landmark has been defined in a consistent manner (13, 17, 39). However, this is a challenging task to reproduce accurately over time with different operators because the three-dimensional structure of individual germaria can be irregular and a dynamic process is being sampled at random times throughout a 12h cycle of FC recruitment. We therefore describe a detailed protocol (see Methods) to promote consistency for all investigators.

The location of FSCs has previously been inferred by examining a large number of marked FSC lineages and using those with only a single candidate stem cell to determine FSC locations (2, 13). This strategy indicated that FSCs were almost all either immediately anterior to the strong Fas3 border (layer 1) or one cell further anterior (layer 2), with a few FSCs in the next most anterior location (layer 3). The total average number of layer 1 (eight) and layer 2 cells (six), distributed in radially equivalent locations abutting the germarial wall, was consistent with estimates of total FSC numbers derived from (i) measuring the average contribution of a single FSC lineage to all FCs and (ii) counting the number of different multicolored lineages in a single ovariole (in each case, taking steps to ensure capturing the full diversity of all FSC lineages, including the most short-lived) (2, 13). After defining FSC locations in this way, it was then possible to count labeled FSCs in layers 1 and 2 explicitly in numerous studies exploring the effects of altered genotypes, regardless of the number of marked FSCs per germarium (17). Some examples of layer 1 and 2 FSCs and their relationship to the strong Fas3 border are shown in Figure 1. Changes in the measured numbers and behaviors of marked FSCs due to altered genotypes in these locations formed a self-consistent picture of FSC responses to different signaling environments (17). Moreover, measurements of absolute rates of FSC division and conversion to FCs and ECs also matched (40), consistent with maintaining a stable population of about 16 FSCs, mainly in layers 1 and 2.

From RNA *in situ*, *santa maria* is mostly expressed posterior to the strong Fas3 border, with weaker expression just anterior to the Fas3 border, overlapping with the posterior margin of *bnb* RNA (37). *Santa maria* RNA is prevalent in group 4 but almost completely absent from group 1, while *bnb* 4 RNA is highest in group 1 but present in some cells of group 4, suggesting that the key Fas3 border is roughly between groups 1 and 4, with most FSCs in group 1 (Fig. 2F, G). A few of the most posterior FSCs, expressing both *bnb* and *santa-maria*, may be in group 4.

*bnb* expression was strongest in group 1 but also significant in abundance and prevalence in group 5 (Fig. 2F, G). Since the anterior boundary of *bnb* RNA expression *in situ* appears to extend beyond FSCs into r2a ECs (37), the transition between ECs and FSCs is likely very close to the border between groups 5 and 1. Other genes with sharp declines from group 5 to 1 include *Hf*, *CG34417* and *CG17278*, with the converse pattern for *CG13698*, *CG42797* and *fz2* (Fig. 3). Several RNAs peak within group 1 (*hdc*, *kank*, *fax*, *sls*, *promL*, *sty*), including *fax* (Fig. 3), for which a protein trap shows a maximum centered on the FSC domain (13).

Regarding the more anterior clusters, *NetA* RNA is restricted to within region 1, while the posterior border of *croc* RNA extends into region 2a, declining quite sharply at the anterior border of *bnb* (37). *NetA* expression was limited to group 0, while *croc* RNA was present also in a minority of group 5 cells, suggesting that groups 0 and 5 largely represent region 1 and 2a ECs, respectively (Fig. 2F, G).

Considering only groups 0, 5 and 1 (478 cells), the total number of cells in groups 0 (231, 48%), 5 (62, 13%) and 1 (185, 39%) are roughly in proportion to the average number of r1 ECs (26, 46%), r2a (14, 25%) and FSCs (16, 29%) in an adult germarium (13, 41, 42), suggesting rough equivalence if all cells were captured equally well for sequencing. Thus, group 1 cells appear to correspond largely to FSCs, though it is expected that a few FSCs may also be within group 4 and not all cells in group 1 are necessarily FSCs (some r2a ECs and some FCs) (Fig. 2D).

### **FSC gene expression patterns**

One motivation for scRNA-seq studies is to identify candidate effectors of FSC behavior. The two basic behaviors of FSCs are division and differentiation. Functional studies have identified the strength of Wnt and JAK-STAT signaling pathways, with opposing intersecting

gradients over the FSC domain, as major determinants of FSC position, differentiation to ECs at the anterior or FCs at the posterior (13, 17, 35). Consequently, it is expected that several downstream responders to these graded signals that execute differentiation responses will have an AP graded expression pattern around the FSC domain. Adhesion proteins are likely important effectors of these signals, potentially promoting posterior FSCs joining a nascent FC epithelium, anterior FSCs moving towards ECs or FSCs moving along a gradient of ECM adhesion ligands. We therefore looked for adhesion molecules (and some partners and regulators) with a clear gradient of expression from group 5 through group 1 to group 4 (Fig. 4A, B). Among the molecules with increasing expression from ECs (group 5) to FCs (group 4) were *Fas3*, *DE-cadherin (shg)*, *Plum*, *Tetraspanin 74F (Tsp74F)*, *bazooka (baz)*, and *coracle (cora) (Dystroglycan (Dg) and Dystrophin (Dys) plateau or peak in group 1)*. The opposite pattern was identified for *Fasciclin I (Fas1)*, *echinoid (ed)*, *Neuroglian (Nrg)*, *Tenascin accessory (Ten-a)*, *Tetraspanin 66E (Tsp66E)*, *scab  $\alpha$ -PS3 integrin (scb)*, *Adherens junction protein p120 (p120ctn)*, *Zyxin (Zyx)* and *Lachesin (Lac)*. Among these genes, only the role of DE-cadherin in FSCs has so far been investigated (43). A more comprehensive listing of expression in different clusters of genes with selectively strong expression in clusters 0, 5, 1 and 4 is presented in Supplementary Table 1.

Posterior FSC cycle faster than anterior FSCs, while ECs do not divide at all (13). Graded JAK-STAT activity, declining from the posterior, is one contributor to this patterned proliferation, while Hedgehog (Hh) and Phospho-inositide 3' kinase (PI3K) pathways also provide major stimuli for FSC division (17, 44-46). Most ECs are found in G1, while most posterior FSCs are in G2, indicating limitation of cycling by G2/M restriction in all FSCs with an increasingly severe G1/S restriction further anterior (40). We found that several RNAs associated with DNA replication increased from anterior (group 0 and 5) to posterior (group 4), including *Proliferating Cell Nuclear Antigen (PCNA)*, *deoxyuridine triphosphatase (dUTPase)*, *Ribonucleotide reductase small (RnrS)* and *large (RnrL)* subunit genes. A similar, graded increase in drivers of mitosis was seen, including *fizzy (fzy)*, *polo*, *cyclin B (cycB)* and *cyclin B3 (cycB3)*, though *string cdc25 phosphatase (stg)* had strongly graded expression of opposite polarity (Fig. 4C).

Signaling pathway activities can be indicated by expression of ligands, although their range of movement can be large, by receptors, which are sometimes up-regulated or down-regulated in response to signal and by genes that commonly respond transcriptionally to pathway activity, though often in tissue-specific ways. *hh* RNA was high in groups 0, 5 and 1, consistent with prior detection in anterior germarial cells, supplementing very strong expression in cap cells and terminal filament cells (47, 48). Correspondingly, the universal Hh target gene, *patched* (*ptc*) declines posteriorly from group 1 cell levels (Fig. 4D). *Wnt6* is expressed largely in group 0, consistent with previous evidence of expression only in anterior ECs, while *Wnt4*, which is strongly expressed in all ECs and FSCs, declines from group 0 through 5, 1 and 4 (36, 42, 48). A similar gradient is seen for *notum* and *naked* (*nkd*), which are sometimes positive transcriptional targets of Wnt signaling (49, 50). Those patterns are consistent with the observed anterior to posterior decline of a Wnt pathway reporter over the FSC domain (13, 51). *Epidermal Growth Factor Receptor* (*EGFR*), one of its ligands *vein* (*vn*), and *rhomboid* (*rho*), which processes Spitz EGFR ligand, all decline from group 5 through group 2 (Fig. 4D). Two genes often induced by BMP signaling, *brinker* (*brk*) and *Daughters against dpp* (*Dad*) show a similar decline. BMP signals supplied to GSCs are strictly limited to the anterior of the germarium (52). Activin signaling in the germarium has not been studied extensively but the agonist ligand *dawdle* (*daw*) was preferentially expressed anteriorly, with the converse pattern for the antagonist *Follistatin* (*Fs*) (53). The only other noted ligand with a strong posterior bias was an insulin receptor agonist, *Insulin-like peptide 6* (*Ilp6*). Expression of a set of imaginal disc growth factors (*Imaginal disc growth factor* (*Idgf*) 1, 2, 3, 6) declined from anterior to posterior (Fig. 4D).

### **Prospective FSC labeling for lineage analyses**

The single-candidate method used to identify FSC locations has not commonly been used for other stem cells. Instead, for mouse stem cells the traditional approach has been to find a sufficiently specific marker to label a subset of cells through a recombination event and then determine by lineage analysis if such labeled cells have stem cell properties (2, 54). In practice, there is often no single suitable marker with absolute specificity. Instead, an empirical



threshold of marker gene expression is engineered to initiate a *Cre-loxP* recombination event under specific conditions of sensitivity. Consequently, questions inevitably remain about whether all cells captured by that strategy are stem cells and whether only a subset of stem cells have been captured (2). These limitations are not commonly voiced but are significant.

ScRNA sequencing of germarial cells can provide candidate genes for targeting FSCs. The most common targeting strategy in *Drosophila* uses specific gene regulatory elements to drive yeast GAL4. GAL4 then activates a *UAS-flp* recombinase gene to initiate stable GFP expression in a recombinant cell and its derivatives (“G-trace”) (55). Temperature-sensitive GAL80 is generally included, so that the timing of recombination events can be limited to a chosen time window at the restrictive temperature. The activity of a specific GAL4 transgene does not necessarily reflect normal expression of the parent gene and is normally assayed by including a *UAS-RFP* reporter. According to such RFP patterns, specific *Wnt4-GAL4*, *C587-GAL4* and *fax-GAL4* lines have been reported to be expressed largely or entirely anterior to the strong Fas3 border (56). We found that after at least 3d at the restrictive temperature, RFP expression was mostly anterior to the strong Fas3 border for all three lines, with strongest expression for *Wnt4-GAL4* and weakest for *fax-GAL4* (Fig. 5A-F). All *Wnt4-GAL4* germaria (24/24) included RFP-positive cells anterior to the Fas3 border in locations we have previously designated as layer 1-3 FSCs, compared to 19/27 (70%) for *fax-GAL4*. Most such cells were labeled in each *Wnt4-GAL4* germarium but only an average of 2.3 (44 total) for *fax-GAL4*. In about a third (8/24) of germaria for *Wnt4-GAL4* there was also detectable RFP in cells immediately posterior to the Fas3 border (generally one cell with weak expression) but RFP label was detected posterior to the Fas3 border in very few (3/59) *fax-GAL4* germaria. For *C587-GAL4*, RFP was detected after 7d at 30C in both layer 1-3 locations anterior to the Fas3 border and at least one cell posterior to the border in all cases (10/10). In the same samples, GFP patterns roughly mirrored RFP expression. However, while RFP expression provides an analog measure of GAL4 activity, GFP expression is digital, with permanent expression at a fixed level (driven by a *ubiquitin* promoter) once expression is triggered by a recombination event. Moreover, the threshold for triggering a stochastic recombination event can be quite low compared to generating a clear RFP signal.

Hence, for lineage studies it is imperative to ascertain the initial GFP labeling pattern and not sufficient to rely on RFP expression patterns.

Previous G-trace tests used long periods at the restrictive temperature and reported that many ovarioles later included GFP-labeled FCs for *Wnt4-GAL4* and *C587-GAL4* (56). Those results are broadly consistent with FSCs being located anterior to the strong Fas3 border. However, a similar experiment using *fax-GAL4* was reported to produce very few ovarioles with labeled FCs under normal conditions in well-fed flies (56). This evidence was used to argue that FSCs lie posterior to the strong Fas3 border. We repeated those conditions and found substantially different results. GFP-labeled FCs were seen in 34% (17/50) of ovarioles; each of these ovarioles also included labeled cells in layer 1-3 (with an average of close to 3 labeled layer 1-3 cells per germarium), while only 9/50 included labeled cells immediately posterior to the Fas3 border. These results, in contrast to the previous report (56), are consistent with layer 1-3 cells as a source of FCs and do not appear to be consistent with the suggestion of long-lived FSCs located posterior to the AP border. Under the same conditions, *C587-GAL4* (10/10= 100%) and *Wnt-GAL4* (24/24= 100% even by 3d) produced a greater frequency and density of FC labeling, but these two GAL4 lines are expressed at higher levels and in more cells anterior to the Fas3 border than *fax-GAL4*.

A much better lineage test would be to label single or very few cells in each germaria over a short time period, ensure no further labeling, and then “chase” for a period sufficient to determine if labeled FCs are formed. We experimented with a variety of conditions to achieve this outcome. A significant practical limitation here (and elsewhere (41, 57)) was the capacity of the temperature-sensitive *GAL80* transgenes to silence GAL4 activity at the permissive temperature. Even in the presence of two such transgenes, GFP clones were occasionally initiated for both *Wnt4-GAL4* and *fax-GAL4* for flies raised and aged at 18C. The nature of those GFP labeling patterns was variable but the frequent heavy-labeling of multiple ECs, FSCs and FCs suggested that most recombination events occurred during development (41), when the expression patterns of *Wnt4-GAL4* and *fax-GAL4* are potentially significantly different from those in adults. *C587-GAL4* with three copies of the *ts-GAL80* transgene gave virtually no background GFP-positive clones at 18C. A second experimental limitation concerns timing prior

to determining initial GFP-labeling patterns. For the weakest GAL4 line, *fax-GAL4*, it was necessary to incubate for 10.5h at the restrictive temperature to have relevant cells labeled in at least 20% of ovarioles. Ideally, flies are then returned to 18C for long enough for all recombination events to manifest as detectable GFP signal before assaying a cohort of flies for initial GFP expression patterns. However, waiting too long will allow some initially labeled cells to divide or to move to different locations, including crossing the strong Fas3 boundary. We developed near-optimal timing for each GAL4 line but the compromise between a sufficient induction period and assaying a true time zero was greatest for *fax-GAL4* because it had the weakest expression. Conversely, the third ideal requirement of limiting GAL4 activity and initial GFP marking strictly to cells anterior to the strong Fas3 border was not met by *Wnt4-GAL4* and *C587-GAL4*. An ideal single GAL4 line, with strong and highly-selective expression, may not be attainable because of the inevitable similarities between a layer 1 FSC and the FC that it can become within a few hours.

Using the best conditions we could establish, we conducted quantitative G-trace experiments for all three GAL4 lines. Samples were scored to ascertain initial GFP labeling patterns (d1) and after a further 6d at 18C (d7). Since development is roughly twice as slow at 18C compared to 25C and egg chambers generally bud every 12h at 25C, it is expected that the most immature FCs (“immediate FCs”) will progress out of the germarium (2d) and occupy the 4th or 5th egg chamber, as previously observed directly for a 3d interval at 25C (9). A FSC may simply remain as a FSC but if it or its progeny become a FC, such FCs would be in the 4<sup>th</sup> egg chamber or more anterior locations. Thus, based on our current understanding, we would expect that all marked FCs anterior to egg chamber 4 on d7 are derivatives of initially-marked FSCs (and a few FSC derivatives may also be present in the 4<sup>th</sup> egg chamber).

For *Wnt4-GAL4*, a small number of ovarioles (<15%) at d1 and d7 included strong, widespread GFP labeling and presumably originated during development, as described above. These background clones were discounted and not included in totals because any additional labeling induced at the restrictive temperature in adults could not be scored. At d1, 19/34 germaria had cells labeled in layers 1-3 anterior to the strong Fas3 border (total of 37 such labeled cells; average 1.9), 14/34 had at least one GFP-positive cell immediately posterior to the

Fas3 border (8 of these also had label in layers 1-3) (Fig. 5K, L). At d7, the observed fraction with marked FCs anterior to the 4<sup>th</sup> egg chamber was 37% (11/30) (Fig. 5O). These results are compatible with the marked FCs at d7 originating from either layer 1-3 cells or cells posterior to the Fas3 border. Essentially, *Wnt-GAL4* initially labels cells posterior to the Fas3 border too frequently to provide a decisive test, even though the pattern of *UAS-RFP* labeling is strongly biased towards cells anterior to the Fas3 border.

For *C587-GAL4* at d1, 14/58 germaria had cells labeled in layers 1-3 anterior to the strong Fas3 border (total of 20 such labeled cells; average 1.4), 8/58 (14%) had a GFP cell immediately posterior to the Fas3 border (4 of these also had label in layers 1-3) (Fig. 5I, J). At 7d, 10/41 (24%) of ovarioles included labeled FCs anterior to the 4<sup>th</sup> egg chamber and a further 5/41 (12%) included labeled cells in layers 1-3 with no labeled FCs (Fig. 5N). The proportion of 7d ovarioles with labeled FCs (24%) exceeded the proportion with labeled cells posterior to Fas3 on d1 (14%), suggesting that there must be a major source of FCs anterior to Fas3. By contrast, the proportion of germaria with labeled layer 1-3 cells at d1 and labeled FCs at d7 were similar, consistent with the former serving as FSCs.

For *fax-GAL4*, 15/123 ovarioles at d1 and 11/81 ovarioles at d7 showed heavily labeled background GFP signal and were excluded. At d1, 26/108 (24%) of germaria had cells labeled in layers 1-3 anterior to the strong Fas3 border (total of 70 such labeled cells; average 2.7), 7/108 (6%) had a GFP cell immediately posterior to the Fas3 border (6 of these also had label in layers 1-3) (Fig. 5G, H). At 7d, 15/70 (21%) of ovarioles included labeled FCs anterior to the 4<sup>th</sup> egg chamber, FCs in the 4<sup>th</sup> egg chamber were the only GFP-positive cells other than ECs in 4/70 (6%) ovarioles, and a further 4/70 (6%) included labeled cells in layers 1-3 with no labeled FCs (Fig. 5M). The proportion of 7d ovarioles with labeled FCs (24%) is much higher than the proportion with labeled cells posterior to Fas3 on d1 (6%), showing that the major source of FCs must be anterior to Fas3. The frequency of ovarioles with labeled layer 1-3 cells at d1 (24%) roughly matched the sum of FC derivatives (21%) and layer 1-3 cells with no FC derivative (6%) at d7, consistent with the expected behavior of layer 1-3 cells as FSCs (2, 13). FCs in the 4<sup>th</sup> egg chamber (6%) may have derived from FSCs or cells immediately posterior to the Fas3 border (6%).

## Discussion

Single-cell RNA sequencing has great potential for understanding key regulators of a cell state. A first step in realizing this potential is to relate single-cell profiles to cell identities. For FSCs, the key intermediate is spatial location. Here, we used cells pre-sorted to represent mainly anterior germarial cells (ECs, FSCs and early FCs) and standard bioinformatic analyses to derive clusters guided by similarities. The clusters were notably similar to those of another study (37), facilitating assignment of group locations according to prior RNA *in situ* results for key indicators of those groups. The cluster profiles themselves showed a clear progression consistent with assignments along the AP axis and the placement of a key spatial landmark, the anterior border of strong surface Fas3 protein expression, roughly at the border between two groups (“1” and “4”). The anterior limit of group “1” was placed roughly three cells further anterior by comparison to RNA *in situ*, forming a cluster size of roughly the expected size to represent FSCs relative to the more anterior groups (“0” and “5”) representing ECs. Importantly, this study is the first to assign the FSC group by taking into account the most comprehensive functional investigation of FSC numbers and locations (13). Some other studies sampled the whole ovariole and therefore lacked sufficient FSCs in their samples (29, 30, 56); others did not acknowledge recent developments in understanding FSC biology or simply avoided an FSC designation, focusing instead on other cell types (37, 56, 57). Here, we included images of a variety of GFP-marked FSC lineages to illustrate FSC locations in 2-3 AP layers anterior to the Fas3 border, as determined previously by examining lineages with only a single candidate FSC (13). We also included a protocol for identifying the Fas3 border because other investigators have presented inconsistent and potentially confusing approaches and summaries (56).

We also undertook a complementary approach to verify FSC locations through prospective “G-trace” labeling (55). We obtained results consistent with our prior designation of FSCs occupying 2-3 layers anterior to the Fas3 border and contradicting an assertion that FSCs are posterior to the AP border (56), based principally on using the very same *fax-GAL4* reagent to initiate clones. Crucially, our tests involved carefully controlled conditions that

allowed production and measurement of a suitable number of GFP-labeled cells during a short labeling period, followed by a chase period to ascertain the fate of such cells. Results with *fax-GAL4* and *C587-GAL4* to initiate labeling were consistent with FC-producing cells lying anterior to the Fas3 border, although that conclusion was less numerically robust for *C587-GAL4* because the associated initial GFP-labeling pattern was less specific than for *fax-GAL4*.

The assignment of a group of scRNA expression profiles to FSCs is not precise for a number of reasons, including imperfect and limited depth of RNA profiles, latitude in spatial correspondence according to RNA *in situs*, and the systematic limitation that FSCs become ECs or FCs within a few hours and therefore inevitably have strong similarities to those neighboring derivatives. Nevertheless, the results of this study have the unique virtue that the approximately-defined FSC group of cells is in accord with current understanding of FSC numbers and locations determined by thorough functional analyses (2, 9, 13, 17, 58-60). FSC behavior is guided by a number of external signals but with prominent roles for two graded signaling pathways, Wnt and JAK-STAT (13, 17). RNAs that also have graded AP expression from ECs through to early FCs are therefore of special interest as candidate mediators of spatially appropriate behavior guided by these graded signals. We have highlighted such RNAs encoding potential effectors of cell movements and differentiation or cell division as candidates for further investigation of functional roles and potential transcriptional regulation by Wnt or JAK-STAT pathways.

## Acknowledgments

This work was supported by the National Science Foundation of China 32172467 to JH, NIH RO1 GM079351 to DK, and a SURF undergraduate research fellowship to XT. We thank John Lis (Cornell University) and the Bloomington stock center for provision of genetic reagents, the Developmental Studies Hybridoma Bank (DSHB) for antibodies, FlyBase as an information resource, David Melamed for preparation of Figure 4, and the confocal microscope resource provided by the Dept. of Biological Sciences, Columbia University.

## Materials and Methods

### Single-cell Suspension Preparation and Fluorescence-activated Cell Sorting (FACS)

Female flies of genotype *C587-GAL4 / yw; UAS-CD8-RFP/+* were selected on the day of eclosion and maintained on rich food at 25C together with males for 4-7d. More than 200 pairs of ovaries were dissected in ice cold DPBS (Dulbecco's Phosphate Buffered Saline) solution, and then gently washed twice by DPBS. The ovaries were digested in 700  $\mu$ l solution containing 0.5% trypsin (Sigma, #T4799) and 0.25% collagenase (Invitrogen, #17018-029) at 25C for about 20 min with gentle shaking. The dissociated cells were filtered through a 40-micron cell strainer (Falcon, #352340), and then Centrifuged at 400 g at 4C for 5 min. Supernatant was removed and the cell pellet was resuspended in DPBS with 0.2% BSA. Cell suspensions were sorted using a FACS Aria III sorter MoFlo XDP (Beckman-Coulter) based on RFP signal. Cells from *w<sup>1118</sup>* fly ovaries were used to set the negative fluorescence gate for the RFP panel. RFP-positive cells were sorted into DPBS containing 0.2% BSA, pelleted (400 g for 10 min) and resuspended in DPBS. The LUNA-FL double fluorescent cell counter (Logos Biosystems) was used to count cell number and the live/dead cell ratio. The suspension of single cells, with > 85% live cells, had a final concentration of 500 cells/ $\mu$ l.

### Single cell RNA-sequencing (scRNA-seq)

The sorted cells were loaded on a Chromium Single cell Controller (10 X Genomics) using the Chromium Single-cell 3' Library & Gel Bead kit v3 according to manufacturer's directions. Briefly, the Gel Beads containing the poly-T primer sequence were linked with a cell barcode and UMI (unique molecular identifier) and the cells are wrapped by "oil droplets" to form a GEM (Gel Bead in Emulsion). Cells were lysed in these droplets and reverse-transcribed to form full-length cDNA sequence. After the oil droplets were broken and purified, the cDNA library was PCR amplified and ligated with sequencing primers. cDNA library quantification assays and quality check analysis were performed using an Agilent 2100 Bioanalyzer and Thermo Fisher Qubit Fluorometer. The library samples were then diluted to a 10 nM concentration and sequenced on an Illumina NovaSeq 6000 platform (Illumina, San Diego) for PE150 sequencing

by Annoroad Gene Tech. Co., Ltd (Beijing, China). A total of 377, 489, 707 reads were obtained for the sample, with 290, 154 mean reads per cell.

### **Raw data preprocessing**

The raw sequencing data was processed with alignment, barcode assignment and UMI counting by Cell Ranger (version 6.1.1) using the previously published pipeline (61). The reference index for Cell Ranger was built using the *Drosophila melanogaster* genome (version BDGP6.32) available on the Ensembl genome database.

### **Single cell RNA-seq data analysis**

Cell Ranger's raw gene count matrix was further analyzed using the Seurat (v4.0.2) R package with standard protocols (62). Only genes expressed in more than 3 cells were kept, and only cells that have unique feature counts between 100 and 4,500, and less than 20% mitochondrial counts were selected for further analysis. Gene expression results were log-normalized, and then regressed on the percentage mitochondrial gene content. The top 16 principal components of Principal component analysis (PCA) were used for clustering according to the Elbow Plot (resolution = 0.9). The ECs, FSC and FCs were identified by markers of each cluster. tSNE plots were used to visualize the clustering results. Mean expression values for each cluster were calculated by "AverageExpression" in Seurat. The R package "pheatmap" was used to generate heatmaps and the gene expression levels were scaled by row.

### **MARCM and multicolor FSC lineages**

In a typical MARCM (mosaic analysis with repressible cell marker) experiment, 1-3d old adult *Drosophila melanogaster* females with the appropriate genotype (*yw hs-Flp, UAS-nGFP, tub-GAL4 /yw; act-GAL80 FRT40A / FRT40A; act>CD2>GAL4 / +* or *yw hs-Flp, UAS-nGFP, tub-GAL4 /yw; FRT42D act-GAL80 tub-GAL80 / FRT42D; act>CD2>GAL4 / +* were given a single 30 min heat shock at 37C. Afterwards, flies were incubated at 25C and maintained by frequent passage on normal rich food supplemented by fresh wet yeast before dissection 6 or 12d later. For multicolor lineages the procedure was similar but with multiple heat shocks (13) and the final genotype of



*yw hs-flp / yw; tub-lacZ FRT40A FRT42B / ubi-GFP FRT40A FRT42B His2Av-mRFP*. Adult female ovaries were dissected in PBS, leaving the tip of the abdomen attached, and fixed in 4% paraformaldehyde (Electron Microscopy Sciences, #15710) for 10 minutes at room temperature, covered to prevent bleaching of fluorophores. The ovaries were then washed three times with 1X PBST (PBS with 0.1% Triton X-100 and 0.05% Tween 20) and blocked at room temperature in blocking solution (PBST with 10% normal goat serum) for 30 minutes. Primary antibodies mouse anti-Fasciclin 3 (Developmental Studies Hybridoma Bank, #7G10, 1:250), and rabbit anti-GFP (Invitrogen, #A6455, 1:1000) were added and samples were nutated for an hour at room temperature. The ovaries were then rinsed with 1X PBST three times for 10 minutes each before being incubated with secondary antibodies Alexa Fluor-647 goat anti-mouse (Invitrogen, #A21236, 1:1000) and Alexa Fluor-488 goat anti-Rabbit (Invitrogen, #A11034, 1:1000) at room temperature for 1 hour. The ovaries were then rinsed twice with PBST and once in PBS. Each ovary pair was then broken apart in PBS on a slide and mounted using *DAPI fluoromount-G* mounting medium (Southern BioTech, OB010020).

Ovarioles were imaged with a Zeiss LSM700 or LSM800 confocal microscope, operated in part by the Zeiss ZEN software.

### **Protocol for identifying cell locations relative to landmark of anterior border of strong Fas3 staining.**

A variety of indicators help to ensure reproducible determination of the anterior border of strong Fas3 staining. These include germarium width, identifying the youngest stage 2b cyst and noting cell process locations. Starting with a mid-z-section, identify the youngest stage 2b cyst by the criteria of spanning the germarium and not being at all rounded. Most germaria have only a single candidate 2b cyst with a clearly more rounded stage 3 cyst more posterior. In such germaria, the Fas3 border runs along the posterior surface of the 2b cyst. In other germaria (up to about a quarter), a new 2b cyst has formed as a more posterior 2b cyst just starts to round. Here, the strong Fas3 border lies between the two 2b cysts. The strong Fas3 border can then be followed as a continuous surface through neighboring z-sections. Layer 1 cells immediately anterior to the Fas3 border have strong Fas3 staining on their posterior

surface but weak or incomplete outlining of the anterior surface by Fas3. Labeling in MARCM clones usually allows visualization of FSC processes (even faintly when using a nuclear-targeted GFP marker). Layer 1 cell processes are present along the Fas3 border, whereas layer 2 processes are anterior to the stage 2b cyst. The widest part of the germarium is generally very close to the Fas3 border, with layer 1 FSCs at the widest location in about three-quarters of samples and immediate FCs at that location in the rest.

### **G-trace experiments**

Flies of genotype *UAS-RFP, UAS-flp, ubi>stop>GFP / CyO; tub-GAL80(ts) / TM2* were crossed to (a) *tub-GAL80(ts) / CyO; fax-GAL4 / TM2* or (b) *Wnt-GAL4 / CyO; tub-GAL80(ts) / TM2* or (c) *C587-GAL4; tub-GAL80(ts) / CyO; tub-GAL80(ts) / TM2* flies at 18C to collect experimental progeny with no balancer chromosomes and two (for *fax-GAL4* and *Wnt4-GAL4*) or three (for *C587-GAL4*) copies of *tub-GAL80(ts)*. Young (1-4d old) experimental flies were then shifted to 29C or 30C for various periods and, in some cases, returned to 18C, as described in the Results for specific tests. For the final lineage tests the conditions were 3h at 30C followed by 13h at 18C (*Wnt-GAL4*), 6h at 30C followed by 13h at 18C (*C587GAL4*), or 10.5h at 30C followed by 13h at 18C (*fax-GAL4*) before dissecting for the first time-point, followed by a further 6d at 18C before dissecting another cohort. An equivalent cohort of flies was kept at 18C throughout. In all cases, flies were transferred to fresh food every 2-3d. Dissection, staining, mounting and analyses were as described for MARCM experiments.

## References

1. Clevers H, Watt FM. Defining Adult Stem Cells by Function, not by Phenotype. *Annu Rev Biochem.* 2018;87:1015-27. doi: 10.1146/annurev-biochem-062917-012341. PubMed PMID: 29494240.
2. Kalderon D. Investigating Adult Stem Cells Through Lineage analyses. *Stem Cell Rev Rep.* 2021. Epub 2021/10/23. doi: 10.1007/s12015-021-10282-z. PubMed PMID: 34677818.
3. Post Y, Clevers H. Defining Adult Stem Cell Function at Its Simplest: The Ability to Replace Lost Cells through Mitosis. *Cell Stem Cell.* 2019;25(2):174-83. doi: 10.1016/j.stem.2019.07.002. PubMed PMID: 31374197.
4. Chatzeli L, Simons BD. Tracing the Dynamics of Stem Cell Fate. *Cold Spring Harbor perspectives in biology.* 2020;12(6). Epub 2020/01/15. doi: 10.1101/cshperspect.a036202. PubMed PMID: 31932319; PMCID: PMC7263083.
5. Dray N, Than-Trong E, Bally-Cuif L. Neural stem cell pools in the vertebrate adult brain: Homeostasis from cell-autonomous decisions or community rules? *Bioessays.* 2021;43(3):e2000228. Epub 2020/12/10. doi: 10.1002/bies.202000228. PubMed PMID: 33295062.
6. Goodell MA, Nguyen H, Shroyer N. Somatic stem cell heterogeneity: diversity in the blood, skin and intestinal stem cell compartments. *Nature reviews Molecular cell biology.* 2015;16(5):299-309. doi: 10.1038/nrm3980. PubMed PMID: 25907613.
7. Jones PH. Stem cell fate in proliferating tissues: equal odds in a game of chance. *Dev Cell.* 2010;19(4):489-90. doi: 10.1016/j.devcel.2010.10.001. PubMed PMID: 20951341.
8. Stine RR, Matunis EL. Stem cell competition: finding balance in the niche. *Trends Cell Biol.* 2013;23(8):357-64. doi: 10.1016/j.tcb.2013.03.001. PubMed PMID: 23597843; PMCID: PMC3729747.
9. Reilein A, Melamed D, Tavaré S, Kalderon D. Division-independent differentiation mandates proliferative competition among stem cells. *Proc Natl Acad Sci U S A.* 2018;115(14):E3182-E91. doi: 10.1073/pnas.1718646115. PubMed PMID: 29555768; PMCID: PMC5889645.
10. Ritsma L, Ellenbroek SI, Zomer A, Snippert HJ, de Sauvage FJ, Simons BD, Clevers H, van Rheenen J. Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live

- imaging. *Nature*. 2014;507(7492):362-5. doi: 10.1038/nature12972. PubMed PMID: 24531760; PMCID: PMC3964820.
11. Rompolas P, Mesa KR, Kawaguchi K, Park S, Gonzalez D, Brown S, Boucher J, Klein AM, Greco V. Spatiotemporal coordination of stem cell commitment during epidermal homeostasis. *Science*. 2016;352(6292):1471-4. doi: 10.1126/science.aaf7012. PubMed PMID: 27229141; PMCID: PMC4958018.
  12. Greulich P, Simons BD. Dynamic heterogeneity as a strategy of stem cell self-renewal. *Proc Natl Acad Sci U S A*. 2016;113(27):7509-14. doi: 10.1073/pnas.1602779113. PubMed PMID: 27313213; PMCID: PMC4941513.
  13. Reilein A, Melamed D, Park KS, Berg A, Cimetta E, Tandon N, Vunjak-Novakovic G, Finkelstein S, Kalderon D. Alternative direct stem cell derivatives defined by stem cell location and graded Wnt signalling. *Nat Cell Biol*. 2017;19(5):433-44. Epub 2017/04/18. doi: 10.1038/ncb3505. PubMed PMID: 28414313; PMCID: PMC5672635.
  14. Beumer J, Clevers H. Cell fate specification and differentiation in the adult mammalian intestine. *Nature reviews Molecular cell biology*. 2020. Epub 2020/09/23. doi: 10.1038/s41580-020-0278-0. PubMed PMID: 32958874.
  15. Donati G, Watt FM. Stem cell heterogeneity and plasticity in epithelia. *Cell Stem Cell*. 2015;16(5):465-76. doi: 10.1016/j.stem.2015.04.014. PubMed PMID: 25957902.
  16. Greenspan LJ, de Cuevas M, Le KH, Viveiros JM, Matunis EL. Activation of the EGFR/MAPK pathway drives transdifferentiation of quiescent niche cells to stem cells in the *Drosophila* testis niche. *Elife*. 2022;11. Epub 2022/04/26. doi: 10.7554/eLife.70810. PubMed PMID: 35468055; PMCID: PMC9038189.
  17. Melamed D, Kalderon D. Opposing JAK-STAT and Wnt signaling gradients define a stem cell domain by regulating differentiation at two borders. *Elife*. 2020;9. Epub 2020/11/03. doi: 10.7554/eLife.61204. PubMed PMID: 33135631; PMCID: PMC7695452.
  18. Rust K, Nystul T. Signal transduction in the early *Drosophila* follicle stem cell lineage. *Curr Opin Insect Sci*. 2020;37:39-48. doi: 10.1016/j.cois.2019.11.005. PubMed PMID: 32087562.
  19. Bastock R, St Johnston D. *Drosophila* oogenesis. *Curr Biol*. 2008;18(23):R1082-7. Epub 2008/12/17. doi: 10.1016/j.cub.2008.09.011. PubMed PMID: 19081037.

20. Eliazer S, Buszczak M. Finding a niche: studies from the *Drosophila* ovary. *Stem cell research & therapy*. 2011;2(6):45. Epub 2011/11/29. doi: 10.1186/scrt86. PubMed PMID: 22117545; PMCID: 3340554.
21. Spradling A, Fuller MT, Braun RE, Yoshida S. Germline stem cells. *Cold Spring Harbor perspectives in biology*. 2011;3(11):a002642. Epub 2011/07/28. doi: 10.1101/cshperspect.a002642. PubMed PMID: 21791699.
22. Decotto E, Spradling AC. The *Drosophila* ovarian and testis stem cell niches: similar somatic stem cells and signals. *Developmental Cell*. 2005;9(4):501-10.
23. Kirilly D, Wang S, Xie T. Self-maintained escort cells form a germline stem cell differentiation niche. *Development*. 2011;138(23):5087-97. Epub 2011/10/28. doi: 10.1242/dev.067850. PubMed PMID: 22031542; PMCID: 3210492.
24. Dai W, Peterson A, Kenney T, Burrous H, Montell DJ. Quantitative microscopy of the *Drosophila* ovary shows multiple niche signals specify progenitor cell fate. *Nature communications*. 2017;8(1):1244. doi: 10.1038/s41467-017-01322-9. PubMed PMID: 29093440; PMCID: PMC5665863.
25. Margolis J, Spradling A. Identification and behavior of epithelial stem cells in the *Drosophila* ovary. *Development*. 1995;121(11):3797-807.
26. Tworoger M, Larkin MK, Bryant Z, Ruohola-Baker H. Mosaic analysis in the *drosophila* ovary reveals a common hedgehog-inducible precursor stage for stalk and polar cells. *Genetics*. 1999;151(2):739-48. Epub 1999/02/02. PubMed PMID: 9927465.
27. Duhart JC, Parsons TT, Raftery LA. The repertoire of epithelial morphogenesis on display: Progressive elaboration of *Drosophila* egg structure. *Mech Dev*. 2017;148:18-39. doi: 10.1016/j.mod.2017.04.002. PubMed PMID: 28433748.
28. Riechmann V, Ephrussi A. Axis formation during *Drosophila* oogenesis. *Curr Opin Genet Dev*. 2001;11(4):374-83. Epub 2001/07/13. doi: S0959-437X(00)00207-0 [pii]. PubMed PMID: 11448623.
29. Jevitt A, Chatterjee D, Xie G, Wang XF, Otwell T, Huang YC, Deng WM. A single-cell atlas of adult *Drosophila* ovary identifies transcriptional programs and somatic cell lineage

- regulating oogenesis. *PLoS biology*. 2020;18(4):e3000538. Epub 2020/04/28. doi: 10.1371/journal.pbio.3000538. PubMed PMID: 32339165; PMCID: PMC7205450.
30. Slaidina M, Gupta S, Banisch TU, Lehmann R. A single-cell atlas reveals unanticipated cell type complexity in *Drosophila* ovaries. *Genome Res*. 2021;31(10):1938-51. Epub 2021/08/15. doi: 10.1101/gr.274340.120. PubMed PMID: 34389661; PMCID: PMC8494228.
31. Forbes AJ, Spradling AC, Ingham PW, Lin H. The role of segment polarity genes during early oogenesis in *Drosophila*. *Development*. 1996;122(10):3283-94.
32. Godt D, Laski FA. Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of *bric a brac*. *Development*. 1995;121(1):173-87.
33. Xie T, Spradling AC. *decapentaplegic* is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell*. 1998;94(2):251-60.
34. Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, Skora AD, Nystul TG, Ohlstein B, Allen A, Wilhelm JE, Murphy TD, Levis RW, Matunis E, Srivali N, Hoskins RA, Spradling AC. The carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics*. 2007;175(3):1505-31. PubMed PMID: 17194782.
35. Vied C, Reilein A, Field NS, Kalderon D. Regulation of stem cells by intersecting gradients of long-range niche signals. *Dev Cell*. 2012;23(4):836-48. Epub 2012/10/20. doi: 10.1016/j.devcel.2012.09.010. PubMed PMID: 23079600; PMCID: 3479678.
36. Waghmare I, Page-McCaw A. Wnt Signaling in Stem Cell Maintenance and Differentiation in the *Drosophila* Germarium. *Genes (Basel)*. 2018;9(3). doi: 10.3390/genes9030127. PubMed PMID: 29495453; PMCID: PMC5867848.
37. Tu R, Duan B, Song X, Chen S, Scott A, Hall K, Blanck J, DeGraffenreid D, Li H, Perera A, Haug J, Xie T. Multiple Niche Compartments Orchestrate Stepwise Germline Stem Cell Progeny Differentiation. *Curr Biol*. 2021;31(4):827-39 e3. Epub 2020/12/29. doi: 10.1016/j.cub.2020.12.024. PubMed PMID: 33357404; PMCID: PMC7904662.
38. McGregor JR, Xi R, Harrison DA. JAK signaling is somatically required for follicle cell differentiation in *Drosophila*. *Development*. 2002;129(3):705-17. Epub 2002/02/07. PubMed PMID: 11830571.

39. Zhang Y, Kalderon D. Hedgehog acts as a somatic stem cell factor in the *Drosophila* ovary. *Nature*. 2001;410(6828):599-604.
40. Melamed D, Choi A, Reilein A, Kalderon D. Spatial regulation of *Drosophila* ovarian Follicle Stem Cell division rates and cell cycle transitions. *bioRxiv*. 2022.
41. Reilein A, Kogan HV, Misner R, Park KS, Kalderon D. Adult stem cells and niche cells segregate gradually from common precursors that build the adult *Drosophila* ovary during pupal development. *Elife*. 2021;10. Epub 2021/10/01. doi: 10.7554/eLife.69749. PubMed PMID: 34590579.
42. Wang X, Page-McCaw A. Wnt6 maintains anterior escort cells as an integral component of the germline stem cell niche. *Development*. 2018;145(3). doi: 10.1242/dev.158527. PubMed PMID: 29361569; PMCID: PMC5818006.
43. Song X, Xie T. DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the *Drosophila* ovary. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(23):14813-8.
44. Hartman TR, Zinshteyn D, Schofield HK, Nicolas E, Okada A, O'Reilly AM. *Drosophila* Boi limits Hedgehog levels to suppress follicle stem cell proliferation. *J Cell Biol*. 2010;191(5):943-52. Epub 2010/11/26. doi: jcb.201007142 [pii] 10.1083/jcb.201007142. PubMed PMID: 21098113; PMCID: 2995164.
45. Huang J, Kalderon D. Coupling of Hedgehog and Hippo pathways promotes stem cell maintenance by stimulating proliferation. *J Cell Biol*. 2014;205(3):325-38. doi: 10.1083/jcb.201309141. PubMed PMID: 24798736; PMCID: PMC4018789.
46. Wang ZA, Huang J, Kalderon D. *Drosophila* follicle stem cells are regulated by proliferation and niche adhesion as well as mitochondria and ROS. *Nature communications*. 2012;3:769. Epub 2012/04/05. doi: 10.1038/ncomms1765. PubMed PMID: 22473013.
47. Forbes AJ, Lin H, Ingham PW, Spradling AC. hedgehog is required for the proliferation and specification of ovarian somatic cells prior to egg chamber formation in *Drosophila*. *Development*. 1996;122(4):1125-35.

48. Sahai-Hernandez P, Nystul TG. A dynamic population of stromal cells contributes to the follicle stem cell niche in the *Drosophila* ovary. *Development*. 2013;140(22):4490-8. doi: 10.1242/dev.098558. PubMed PMID: 24131631; PMCID: PMC3817939.
49. Kakugawa S, Langton PF, Zebisch M, Howell S, Chang TH, Liu Y, Feizi T, Bineva G, O'Reilly N, Snijders AP, Jones EY, Vincent JP. Notum deacylates Wnt proteins to suppress signalling activity. *Nature*. 2015;519(7542):187-92. Epub 2015/03/04. doi: 10.1038/nature14259. PubMed PMID: 25731175; PMCID: PMC4376489.
50. Zeng W, Wharton KA, Jr., Mack JA, Wang K, Gadbow M, Suyama K, Klein PS, Scott MP. naked cuticle encodes an inducible antagonist of Wnt signalling. *Nature*. 2000;403(6771):789-95. Epub 2000/02/29. doi: 10.1038/35001615. PubMed PMID: 10693810.
51. Wang X, Page-McCaw A. A matrix metalloproteinase mediates long-distance attenuation of stem cell proliferation. *J Cell Biol*. 2014;206(7):923-36. doi: 10.1083/jcb.201403084. PubMed PMID: 25267296; PMCID: PMC4178971.
52. Harris RE, Ashe HL. Cease and desist: modulating short-range Dpp signalling in the stem-cell niche. *EMBO Rep*. 2011;12(6):519-26. doi: 10.1038/embor.2011.80. PubMed PMID: 21546910; PMCID: PMC3128284.
53. Herrera SC, Sainz de la Maza D, Grmai L, Margolis S, Plessel R, Burel M, O'Connor M, Amoyel M, Bach EA. Proliferative stem cells maintain quiescence of their niche by secreting the Activin inhibitor Follistatin. *Dev Cell*. 2021. Epub 2021/08/08. doi: 10.1016/j.devcel.2021.07.010. PubMed PMID: 34363758.
54. Kretzschmar K, Watt FM. Lineage tracing. *Cell*. 2012;148(1-2):33-45. doi: 10.1016/j.cell.2012.01.002. PubMed PMID: 22265400.
55. Evans CJ, Olson JM, Ngo KT, Kim E, Lee NE, Kuoy E, Patananan AN, Sitz D, Tran P, Do MT, Yackle K, Cespedes A, Hartenstein V, Call GB, Banerjee U. G-TRACE: rapid Gal4-based cell lineage analysis in *Drosophila*. *Nature methods*. 2009;6(8):603-5. doi: 10.1038/nmeth.1356. PubMed PMID: 19633663; PMCID: PMC2754220.
56. Rust K, Byrnes LE, Yu KS, Park JS, Sneddon JB, Tward AD, Nystul TG. A single-cell atlas and lineage analysis of the adult *Drosophila* ovary. *Nature communications*. 2020;11(1):5628.



Epub 2020/11/08. doi: 10.1038/s41467-020-19361-0. PubMed PMID: 33159074; PMCID: PMC7648648.

57. Shi J, Jin Z, Yu Y, Zhang Y, Yang F, Huang H, Cai T, Xi R. A Progressive Somatic Cell Niche Regulates Germline Cyst Differentiation in the *Drosophila* Ovary. *Curr Biol*. 2021;31(4):840-52 e5. Epub 2020/12/20. doi: 10.1016/j.cub.2020.11.053. PubMed PMID: 33340458.
58. Fadiga J, Nystul TG. The follicle epithelium in the *Drosophila* ovary is maintained by a small number of stem cells. *Elife*. 2019;8. doi: 10.7554/eLife.49050. PubMed PMID: 31850843; PMCID: PMC6946398.
59. Kalderon D, Melamed D, Reilein A. Follicle Stem Cells (FSCs) in the *Drosophila* ovary; a critique of published studies defining the number, location and behavior of FSCs. *bioRxiv*. 2020:2020.06.25.171579. doi: 10.1101/2020.06.25.171579.
60. Kalderon D, Melamed D, Reilein A. The number of Follicle Stem Cells in a *Drosophila* ovariole. *bioRxiv*. 2021:2021.10.25.465475 doi: 10.1101/2021.10.25.465475
61. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. Massively parallel digital transcriptional profiling of single cells. *Nature communications*. 2017;8:14049. Epub 2017/01/17. doi: 10.1038/ncomms14049. PubMed PMID: 28091601; PMCID: PMC5241818 L.M., D.A.M., S.Y.N., M.S.L., P.W.W., C.M.H., R.B., A.W., K.D.N., T.S.M. and B.J.H. are employees of 10x Genomics.
62. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-87 e29. Epub 2021/06/02. doi: 10.1016/j.cell.2021.04.048. PubMed PMID: 34062119; PMCID: PMC8238499.

## Figure Legends

### Figure 1. Location of FSCs in the *Drosophila* germarium.

(A) Germline Stem Cell (GSC) daughters (gray) develop into 16-cell cysts as they progress further posterior (right) in region 1, supported by interactions with quiescent enveloping Escort Cells (ECs) and span the width of the germarium in region 2b as ECs are replaced by Follicle Cells (FCs). Two potential marked FSC lineages are shown (blue, green), together with the pattern of strong Fas3 surface protein expression (red). (B-J) Examples of FSC lineages marked by positive marking and multicolor lineage tracing. The anterior border of strong Fas3 generally aligns with the posterior of the stage 2b germline cyst, as in (A), and is indicated by broken white lines in (C', D', E', F', G', H-J); central z-sections are shown but experiments examine cells in all z-sections. (B-G) show GFP-positive (green) MARCM FSC lineages with Fas3 (red) and (C-E) Vasa (blue) marking germline cells. FSC nuclei are indicated by colored arrows according to AP location of layer 1 (white), layer 2 (cyan) or layer 3 (yellow), from posterior to anterior. (B) Most FSC lineages that survive for several days contain several FSCs (arrows) and one or more labeled EC (arrowhead), precluding identification of which of these cells may be maintaining the lineage. Lineages with only a single candidate stem cell (C, D, F, G) or with candidates in only a single AP plane (E) were comprehensively scored to reveal that an FSC could be found immediately anterior to the Fas3 border (white arrows), one cell further anterior (blue arrows) or occasionally one cell further anterior. (H-K) Multicolor FSC lineages, marked by the loss of GFP (G, green), RFP (R, red) or lacZ (B, blue) in different combinations were analyzed in the same way to reveal single candidates or single-plane candidates, as indicated by colored arrows (derivative FCs of each are indicated by arrowheads of matching color). Single candidate FSCs were also found to be evenly distributed around the germarial circumference (in all z-sections). (K) A germarium oriented perpendicular to others shows that in cross-section around the AP plane of layer 1 FSCs there are multiple somatic cell nuclei (nine here, labeled according to retained colors) surrounding a central germline cyst. Scale bar 20  $\mu\text{m}$  in (B-J), 10 $\mu\text{m}$  in (K).

## Figure 2. Cluster analysis of scRNA-seq profiles of germarial cells.

(A) t-Distributed Stochastic Neighbor Embedding (t-SNE) plot of the dataset, together with (C) the number of unique molecular identifiers (nUMI) detected per cell and percentage of mitochondrial gene for each cluster. (B) t-SNE plot after re-clustering the original groups 1, 3, 5, 2 only. (D) Deduced location in the germarium of cells within the six clusters shown in (C). (E) Summary dot plot of key marker genes defining clusters presented in (37), together with dot plots of the same genes for the six clusters in (C). Double-headed arrows indicated similar groups in the two studies (groups 0 and 5 span the content of groups IGS1-3; *phm* was not detected in our study). (G) Expression of the six most informative marker genes imposed on t-SNE plot of (C) groups 0-5.

## Figure 3. Marker genes for six clusters of somatic germarial cells.

Dot plots showing the expression profiles of the top 30 markers (ordered by Adjusted p-value) in each of the clusters, (A) cluster 0, (B) cluster 5, (C) cluster 1, (D) cluster 4, (E) cluster 2 and (F) cluster 3, proceeding from anterior to posterior cell groups in the germarium.

## Figure 4. Potential regulators of cell behavior with AP-graded expression patterns.

Heatmaps of selected genes encoding (A) adhesion molecules, (B) adhesion-related molecules, (C) drivers or regulators of cell cycling and (D) selected signaling molecules with pronounced differences in expression from ECS (groups 0 and 5), through FSCs (group 1) to early FCs (group 2).

## Figure 5. G-trace lineage analysis with *fax-GAL4*, *C587-GAL4* and *Wnt-GAL4*.

(A-F) Expression of *UAS-RFP* after incubation at the restrictive temperature for (A, B, E, F) 3d or (C, D) 7d, with Fas3 antibody staining (white). In all cases, several ECs, far anterior (left) to strong Fas3 and arrows, express RFP (red). (A, B) RFP was seen in some (A) layer 1 (white arrow) and layer 2 (blue arrow) cells but (B) not in all samples for *fax-GAL4*. (C, D) RFP was seen in layer 1 (white arrows), layer 2 (blue arrows) and layer 3 (yellow arrows), as well as more faintly in

cells just posterior to the strong Fas3 border (pink arrows) for *C587-GAL4*. (E, F) RFP was seen in layer 1 (white arrows), layer 2 (blue arrows) and layer 3 (yellow arrows), as well as more faintly in cells just posterior to the strong Fas3 border (pink arrows) in (E) some samples but (F) not others for *Wnt4-GAL4*. (G-L) GFP (green) expression at d1 after incubation at the restrictive temperature (30C) and return to 18C. (G, H) For *fax-GAL4*, GFP was common in ECs, occasionally present in layer 1, 2 (blue arrow) or 3 (yellow arrows) cells (24%) and rarely present posterior to the strong Fas3 (white) border (6%). (I, J) For *C587-GAL4*, GFP was detected in ECs, sometimes in layer 1, 2 (blue arrows) or 3 cells (24%) and less often (J) immediately posterior to the strong Fas3 border (pink arrow) (14%). (K, L) For *Wnt4-GAL4*, GFP was detected in ECs, frequently in layer 1, 2 or 3 (yellow arrow) cells (56%) and almost as often (L) immediately posterior to the strong Fas3 border (pink arrow) (41%). (M-O) GFP (green) expression after a further 6d at 18C was found in small groups of FCs in the germarium or first three egg chambers in a subset of samples for (M) *fax-GAL4* (21%), (N) *C587-GAL4* (24%) and (O) *Wnt4-GAL4* (37%). Scale bar of (A) applies also to (B-N); all are 20 $\mu$ m.

## Supplementary Materials

**Figure S1. Identity markers for peripheral groups in initial t-SNE clusters.** Related to Figure 2. (A-C) t-SNE plots of complete data set (Fig. 2A) showing color-coded relative expression levels of genes characteristic of (A) TF or cap cells (*Lmx1a*, *en*, *dpp*, *wg*), (B) stalk or pre-stalk cells (*LamC*, *sim*) or (C) germline cells (*zpg*, *chinmo*, *ovo* and *otu*).

### Supplementary Table 1

Ranked list of genes, ordered by adjusted p-value ( $p_{val\_adj}$ ), for each cluster (0, 5, 1, 4, 2, 3), showing the Log fold-change of the average expression between clusters and the percentage of cells where the gene is detected in the cluster. Clusters correspond to Fig. 2C.

### Supplementary Table 2

Genes with *increasing* expression from cluster 0 through clusters 5 and 1 to cluster 4. These genes were picked by calculating the Spearman correlation coefficient to find genes with a positive coefficient greater than 0.4. Average expression value for each cluster (including clusters 2 and 3) is listed (“AverageExpression”, described in Methods). The change from cluster 0 to cluster 4 (“updegree”) is calculated as  $(cluster4 - cluster0)/cluster4$ . The resulting values were classified into quartiles, where the lowest number represents the greatest variation between groups 0 and 4. Genes are ordered according to the absolute expression value in cluster 4.

### Supplementary Table 3

Genes with *decreasing* expression from cluster 0 through clusters 5 and 1 to cluster 4. These genes were picked by calculating the Spearman correlation coefficient to find genes with a negative coefficient greater than 0.4 magnitude. Average expression value for each cluster (including clusters 2 and 3) is listed (“AverageExpression”, described in Methods). The change from cluster 0 to cluster 4 (“updegree”) is calculated as  $(cluster0 - cluster4)/cluster4$ . The resulting values were classified into quartiles, where the lowest number represents the greatest variation between groups 0 and 4. Genes are ordered according to the absolute expression value in cluster 0.

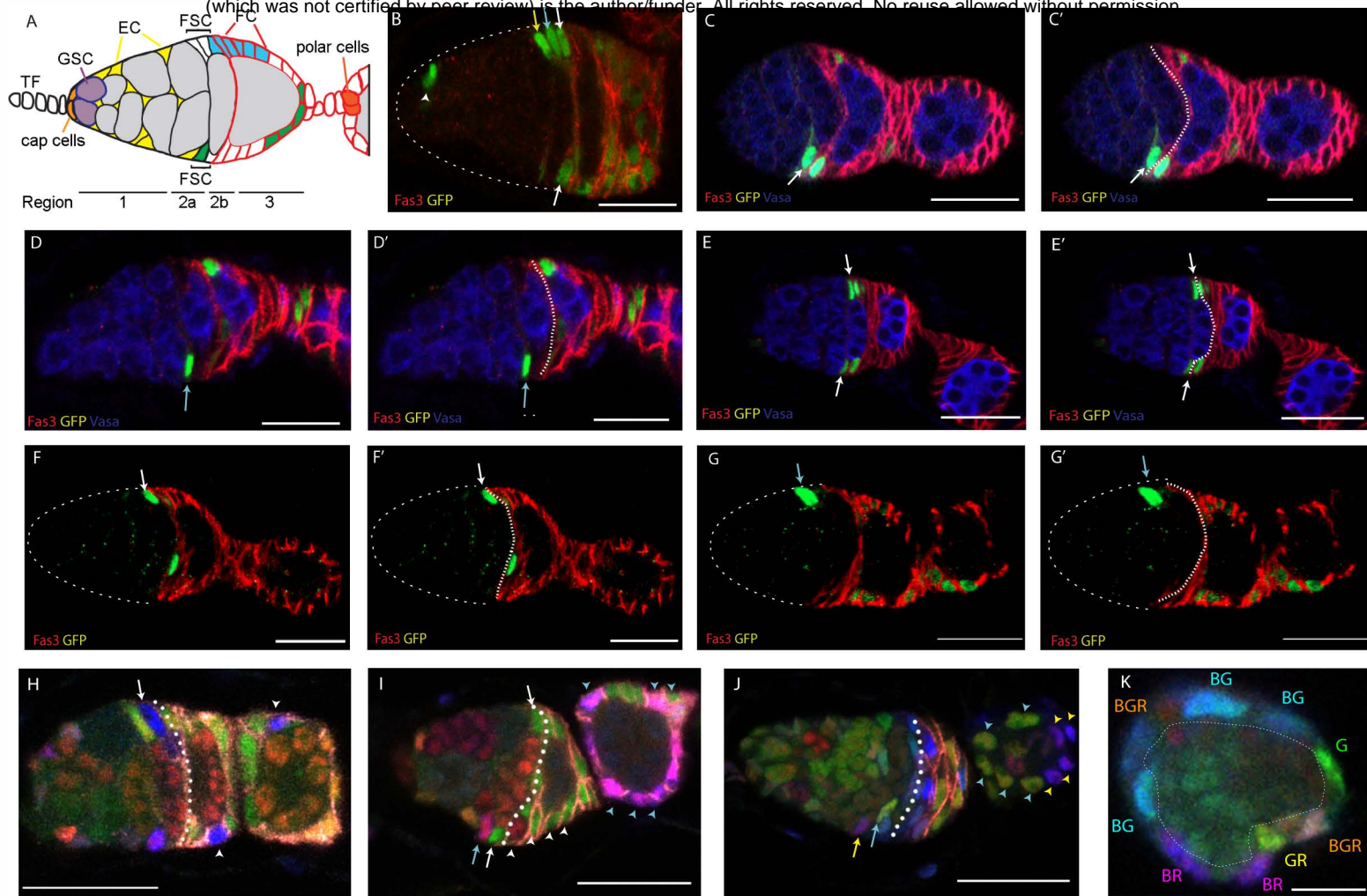


Figure 1



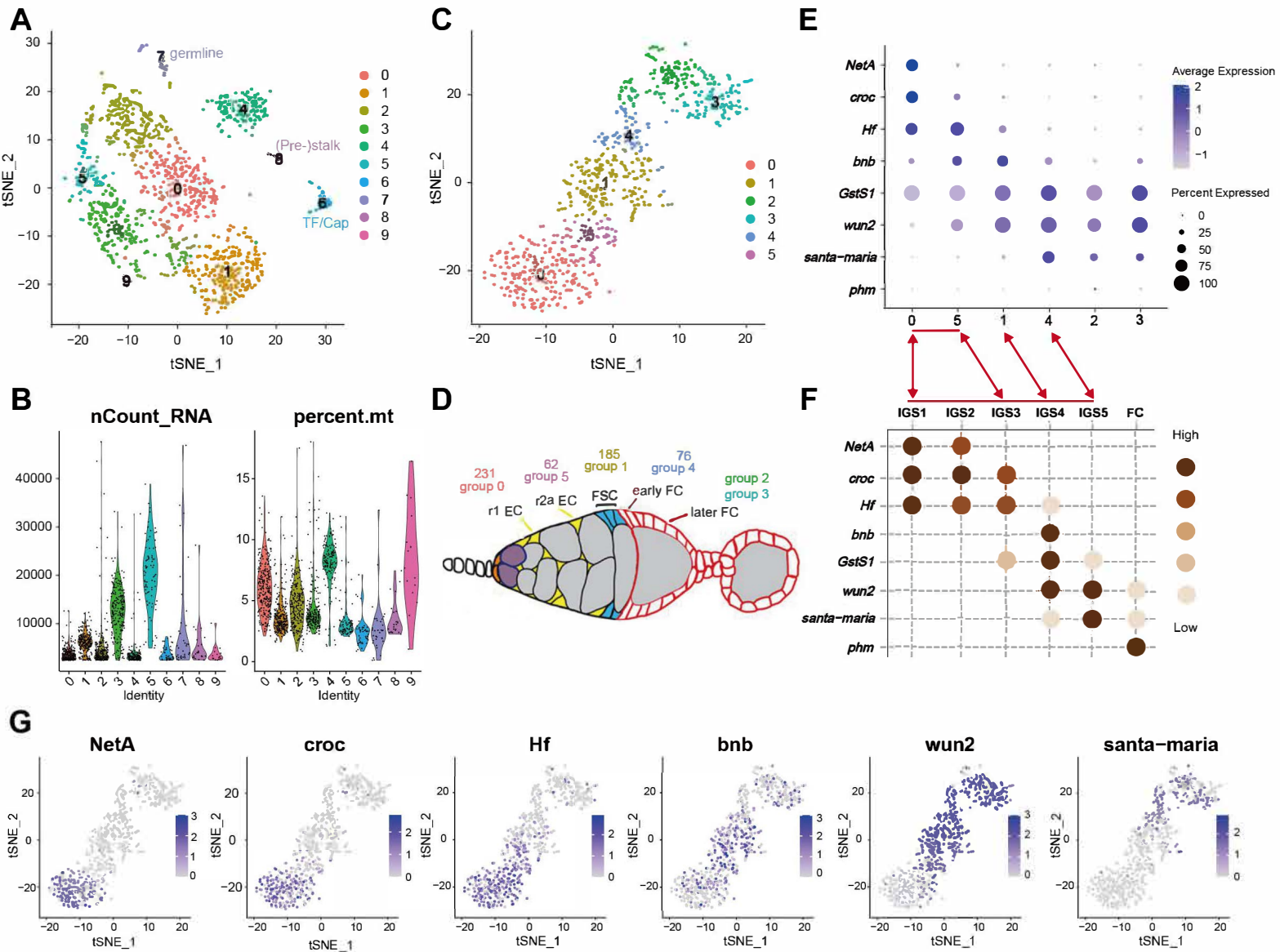


Figure 2

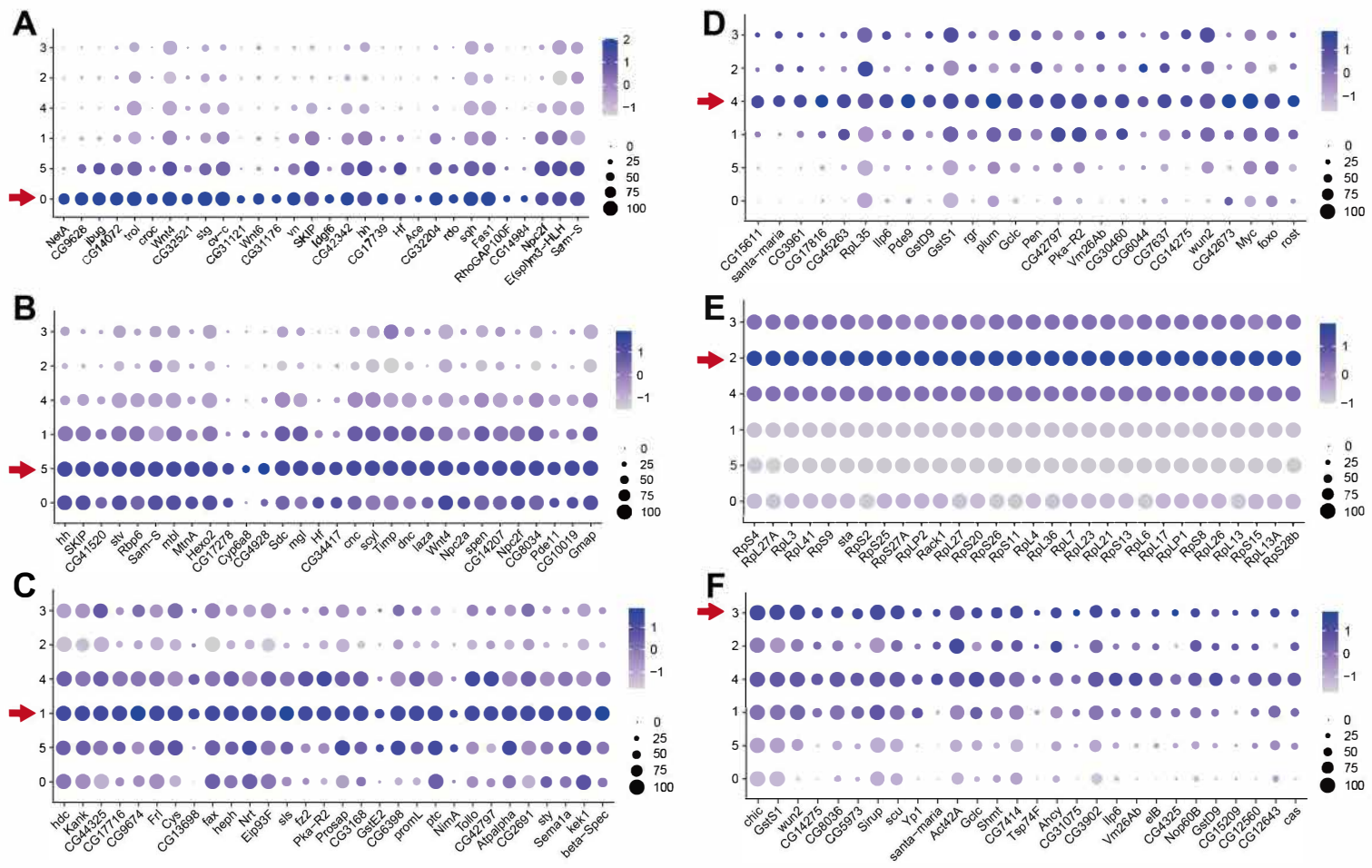


Figure 3



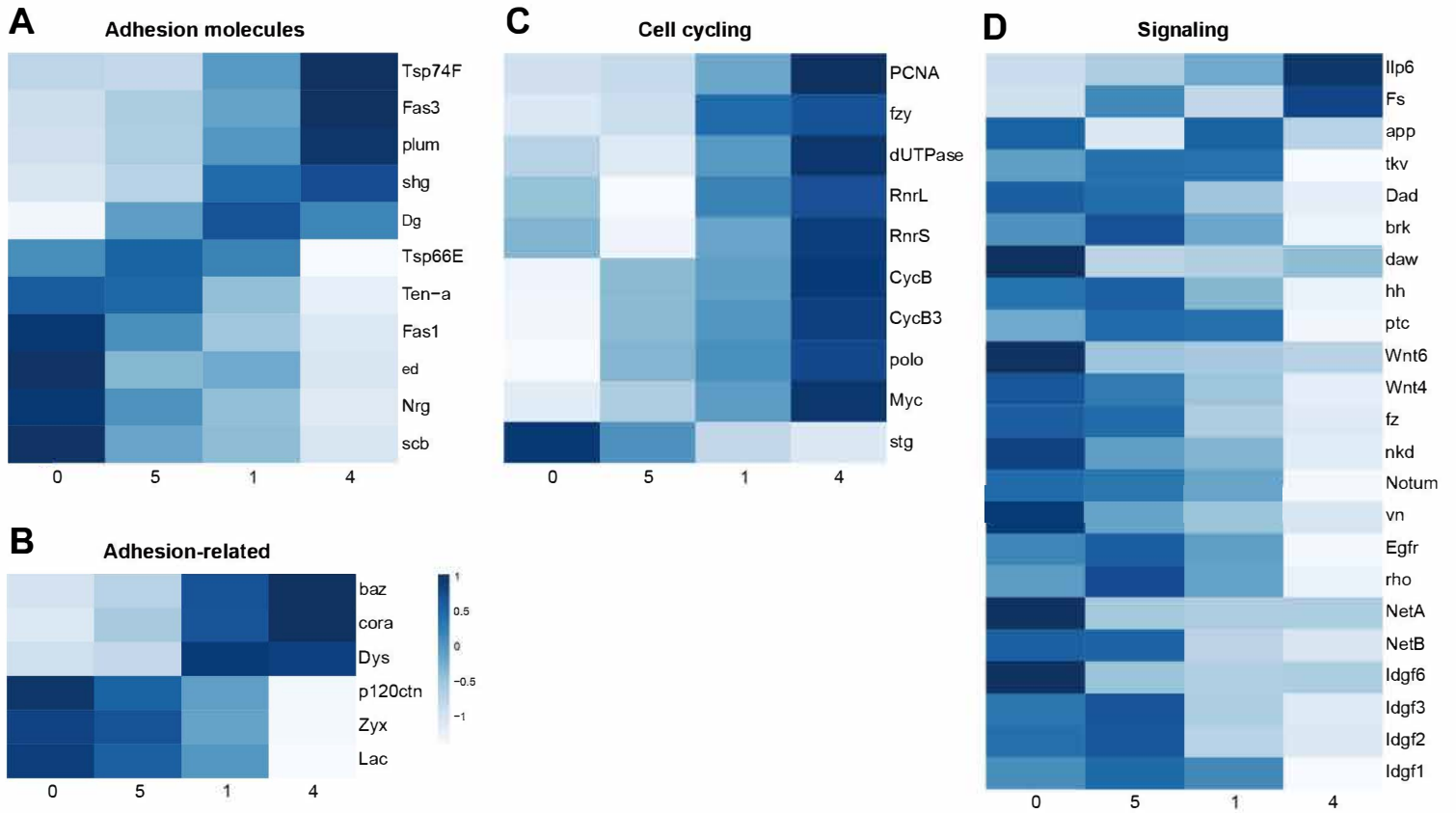


Figure 4

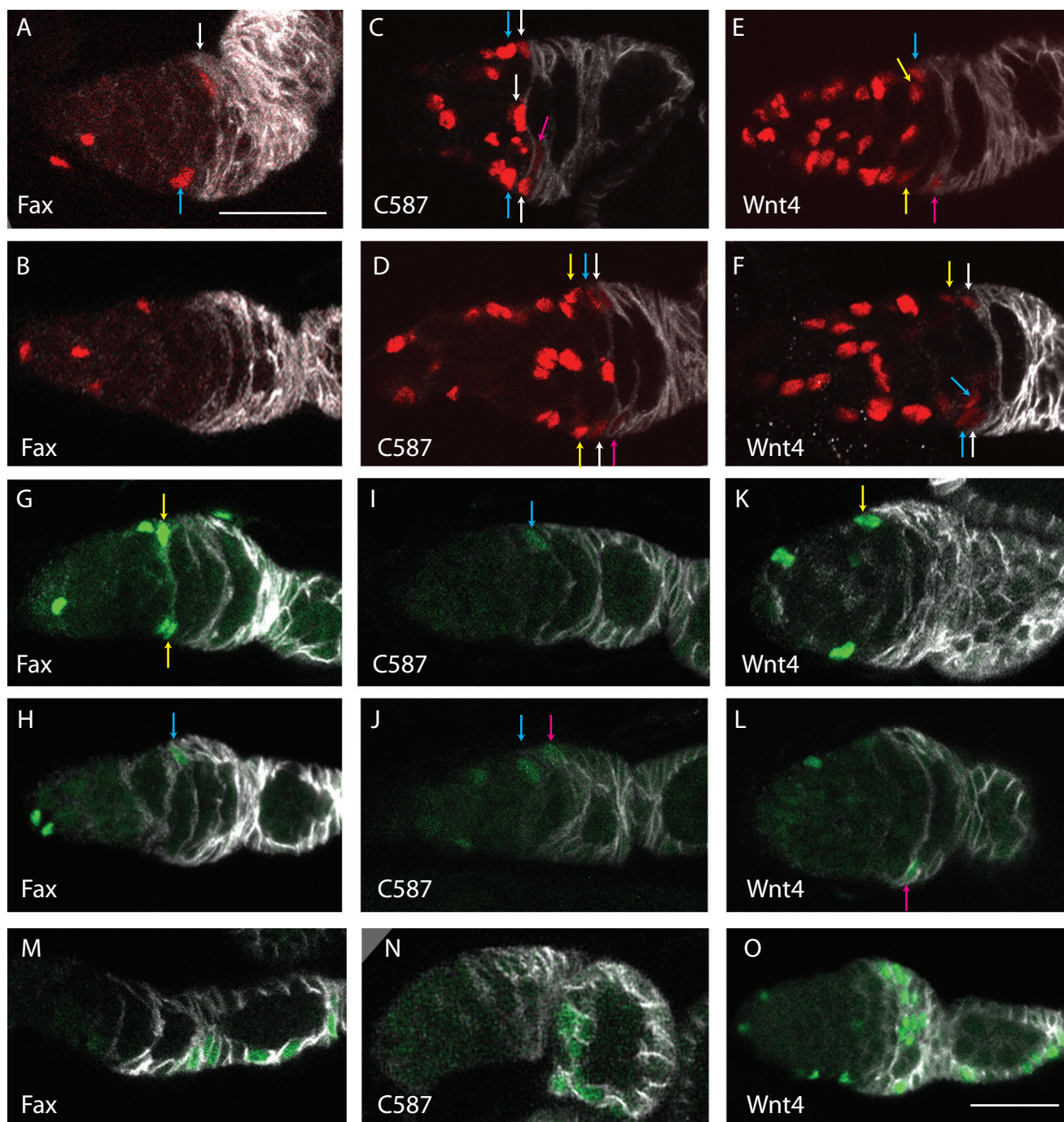


Figure 5