Distinct gene expression dynamics in germ line and somatic tissue during ovariole morphogenesis in *Drosophila melanogaster*

Shreeharsha Tarikere[1a], Guillem Ylla[1a] and Cassandra G. Extavour[1, 2*]

1. Department of Organismic and Evolutionary Biology, Harvard University
2. Department of Molecular and Cellular Biology, Harvard University

* Author for correspondence: extavour@oeb.harvard.edu

a. These authors contributed equally to this work

Abstract

The survival and evolution of a species is a function of the number of offspring it can produce. In insects the number of eggs that an ovary can produce is a major determinant of reproductive capacity. Insect ovaries are made up of tubular egg-producing subunits called ovarioles, whose number largely determines the number of eggs that can be potentially laid. Ovariole number is directly determined by the number of cellular structures called terminal filaments, which are stacks of cells that assemble in the larval ovary. Elucidating the developmental and regulatory mechanisms of terminal filament formation is thus key to understanding the regulation of insect reproduction through ovariole number regulation. We systematically measured mRNA expression of all cells in the larval ovary at the beginning, middle and end of terminal filament formation. We also separated somatic and germ line cells during these stages and assessed their tissue-specific gene expression during larval ovary development. We found that the number of differentially expressed somatic genes is highest during late stages of terminal filament formation and includes many signaling pathways that govern ovary development. We also show that germ line tissue, in contrast, shows greater differential expression during early stages of terminal filament formation, and highly expressed germ line genes at these stages largely control cell division and DNA repair. We provide a tissue-specific and temporal transcriptomic dataset of gene expression in the developing larval ovary as a resource to study insect reproduction.

KEY WORDS: Ovary, FACS, RNA-Seq, Terminal filament, Germ line, Stem cell niche.

INTRODUCTION

Healthy reproductive organs are among the most important factors that determine the fertility of an individual, and more importantly, continuity of the species itself. Reproductive fitness, including fecundity, is determined by the number of progenies an organism can produce. In insects, egg-producing subunits of ovaries are called ovarioles (BÜNING 1994). In flies of the genus *Drosophila*, the number of ovarioles predicts the peak egg laying potential of the females of the species (DAVID 1970), and is negatively correlated with egg size but positively correlated with reproductive output (CHURCH *et al.* 2021). The number of ovarioles varies widely across insects and is in the range of 18-24 ovarioles per ovary in wild type North American populations of *Drosophila melanogaster* (HONEK 1993; MARKOW AND O'GRADY 2007; HODIN 2009). In *Drosophila*, adult ovariole number is established in the larval stages through the development of a species-specific number of linear somatic cell stacks called terminal filaments (KING *et al.* 1968).The number of terminal filaments assembled by the time of pupariation usually predicts adult ovariole number (KING 1970; HODIN AND RIDDIFORD 2000a). Thus, terminal filaments are the primordial larval structures whose number ultimately determines ovariole number. The genetic mechanisms governing ovary morphogenesis, which includes the process of regulation of terminal filament number and assembly during larval ovary development, remain poorly understood.

Ovary morphogenesis is orchestrated by interactions of the cell types of somatic and germ line tissues. Larval somatic ovarian tissue is principally made up of five cell types - sheath cells, swarm cells, terminal filaments, cap cells, and intermingled cells. The anterior most cells of the ovary are the sheath cells, and a sub-population of these apically

positioned cells undergo two cell migration events during larval ovary development. First, a population of sheath cells called swarm cells migrates from the anterior to the posterior of the ovary to form the basal region in the mid third larval instar stage (COUDERC *et al.* 2002; GREEN II AND EXTAVOUR 2012). Secondly, in the late third instar and early pupal stages, sheath cells migrate from the apical to the basal region, traversing in between terminal filaments (KING *et al.* 1968). These sheath cells lay down basement membrane in their path, which encapsulates developing ovarioles (KING 1970). Terminal filaments are stacks of cells located just below the sheath cells in the anterior larval ovary. They are formed by a process of progressive intercalation of flattened cells into stacks, and stack formation occurs in a "wave" that proceeds from the medial to the lateral side in the larval ovary (**Figure 1A**) (SAHUT-BARNOLA *et al.* 1995a).

The genes *bric á brac 1 (bab1) and bric á brac 2 (bab2)* are expressed in the terminal filaments and essential for terminal filament cell differentiation and terminal filament assembly (GODT AND LASKI 1995b; SAHUT-BARNOLA *et al.* 1995a; COUDERC *et al.* 2002; SALER *et al.* 2020).The gene *engrailed* is also expressed in terminal filaments, and at a lower level in the cap cells of the larval ovary (BOLIVAR *et al.* 2006; SALER *et al.* 2020). Clones homozygous for an *engrailed* deletion allele generated in terminal filaments in third instar larvae showed that this gene is required in initial terminal filament precursors for the correct assembly of terminal filaments. (BOLIVAR *et al.* 2006). However, a subsequent study showed that RNAi knockdowns of *engrailed* and *invected* driven by *bab:GAL4* in larval terminal filament and cap cells does not affect terminal filament formation (SALER *et al.* 2020). This could mean that *engrailed/invected* are not absolutely required for terminal filament formation, but that genetic heterogeneity with respect to

engrailed/invected dose, is important among terminal filament precursor cells to ensure correct terminal filament morphogenesis. Accumulation of *engrailed* in terminal filaments is dependent on *bab* gene expression (SALER *et al.* 2020).

We previously showed that the *Hippo signaling* pathway controls the regulation of cell proliferation in somatic cells, thereby affecting the number of terminal filaments and their constituent terminal filament cells (SARIKAYA AND EXTAVOUR 2015). During early terminal filament formation, Actin and Armadillo (arm) proteins deposited in the region between terminal filaments make a scaffold to flatten and intercalate terminal filament cells (GODT AND LASKI 1995b; SAHUT-BARNOLA *et al.* 1995a; CHEN *et al.* 2001). Expression of the protein cofilin (*twinstar*) is required in terminal filament and apical cells for actin-based change in cell shape, and loss of cofilin causes a reduction in terminal filament and apical cell numbers (CHEN *et al.* 2001).

Normal growth of an ovary depends on the homeostatic proliferation of the somatic and germ line tissues (GILBOA AND LEHMANN 2006; GILBOA 2015). This balance between somatic and germ line tissue populations is achieved by regulation of proliferation, differentiation and apoptosis of stem cell populations of somatic and germ cell lineages (SAHUT-BARNOLA *et al.* 1995a; SAHUT-BARNOLA *et al.* 1996). Somatic cells called intermingled cells interact with the germ cells and control their proliferation (LI *et al.* 2003; GILBOA AND LEHMANN 2006; SARIKAYA AND EXTAVOUR 2015; LAI *et al.* 2017; PANCHAL *et al.* 2017; LI *et al.* 2019). *Notch, hedgehog, Mitogen Activated Protein Kinase (MAPK)* and *Epidermal growth factor receptor* (*EGFR)* signaling pathways, as well as the transcription factor *traffic jam* maintain the germ line stem cell niche (BESSE *et al.* 2005; SONG *et al.*

2007; MATSUOKA *et al.* 2013; SARIKAYA AND EXTAVOUR 2015; YATSENKO AND SHCHERBATA 2021), which is established at the base (posterior) of each terminal filament.

Recent work by Slaidina and colleagues used single-cell transcriptomics to describe the gene expression profiles of the various cell types of the late third instar larval ovary (SLAIDINA *et al.* 2020). They sub-divided terminal filament cells into anterior or posterior cell types, and sheath cells into migratory or non-migratory cell types, based on gene expression patterns of the single cell sub-populations. (SLAIDINA *et al.* 2020). While this study examined a single time point of ovary development, given that ovary morphogenesis is a temporal process, we hypothesize that changes in gene expression patterns over the course of development may be important to regulate morphogenesis. Thus, a gene expression study across the developing stages of larval ovary would advance our understanding of the transcriptomic regulation of ovarian morphogenesis.

Although all major conserved animal signaling pathways are known to be involved in ovarian morphogenesis (TWOMBLY *et al.* 1996; COHEN *et al.* 2002; HUANG *et al.* 2005; SONG *et al.* 2007; GANCZ AND GILBOA 2013; GREEN AND EXTAVOUR 2014; SARIKAYA AND EXTAVOUR 2015; KUMAR *et al.* 2020), a systematic gene expression profile of a developing ovary is lacking. Such system-wide gene expression data for the ovary throughout terminal filament morphogenesis, including the potentially distinct transcriptional profiles of germ cells and somatic cells, could shed light on the processes involved in the maintenance of cell types necessary to shape the ovary and control the number of ovarioles.

To this end, we measured gene expression during the development of the larval ovary by systematically staging and sequencing mRNA from whole ovaries before, during

and after terminal filament formation. Furthermore, we separated somatic and germ line tissue types at each of these stages to analyze tissue-specific gene expression. We compared the gene expression profiles across tissues and also across stages of ovary development. We then employed functional enrichment analysis to determine the different biological functions active in the three larval developmental stages and two tissue types that could yield information on ovary morphogenesis. This dataset is an important temporal and tissue specific gene expression resource for the insect developmental biology community to understand early ovary development.

RESULTS

Staging larval ovary development during terminal filament formation

We divided the developing *Drosophila* larval ovary into three stages during terminal filament formation and used RNA-seq to quantify gene expression at these stages (**Figure 1A**). First, we considered an early stage of terminal filament formation at the early third instar larva (72 hours After Egg Laying, 72h AEL), when terminal filament assembly is initiating (GODT AND LASKI 1995b) (**Figure 1A-i)**. Second, we assigned the middle (mid) stage (96h AEL) as 24 hours after the early stage, at the midway point of terminal filament assembly (GODT AND LASKI 1995b) (**Figure 1A-ii**). Third, the late stage (120h AEL) was defined as the time point of white pupa formation (when the larvae become immobile at the larval to pupal transition (ASHBURNER *et al.* 2005)), which occurs 24 hours after the middle stage (**Figure 1A-iii**). At the white pupa stage, terminal filament assembly is complete and the number of terminal filaments reflects the number of adult ovarioles (HODIN AND RIDDIFORD 2000b) (**Figure 1B**).

We dissected these three stages of developing ovaries from larvae obtained from synchronized eggs and sequenced the transcripts present at each stage from pools of 30-100 ovaries **(Supplementary Table S1)**. We aligned reads to the *Drosophila melanogaster* genome (FlyBase v6.36), which yielded between 88.49% and 98.06% of reads aligned per sample (**Supplementary Figure S1, Supplementary Table S1**). Clustering analysis based on the variance-stabilizing transformation (VST) of the gene counts of each sample confirmed that the three biological replicates of each stage clustered together, and that the three stages were well separated, as reflected by the dendrogram of the hierarchical analysis and the principal component analysis (PCA) (**Figure 2A-B**). Furthermore, the dendrogram visualization of the hierarchical clustering results revealed that the mid stage was more similar in expression profile to the early stage than to the late stage. This indicates a more pronounced transcriptomic change at the transition from mid to late, than from early to mid, despite the fact that the same chronological amount of time had elapsed between each stage.

Differential gene expression analysis of whole ovary samples at different stages

We analyzed the transcriptional differences between each stage and the successive one, thus performing a differential expression analysis comparing early to mid and mid to late transitions, using DESeq2 (LOVE *et al.* 2014) with a threshold of $p<0.01$ (see Methods). We found a significantly higher number of genes differentially expressed in the mid to late transition (2,727 genes), than in the early to mid-transition (685) (**Figure 2C, Supplementary Table S2**). Interestingly, from early to mid stages twice as many genes were downregulated (480) as upregulated (205), while from mid to late stages

171 approximately the same proportion of genes were upregulated (1,264) and
172 downregulated (1,463). We then identified the genes that were differentially expressed in
173 one stage as compared to the other two stages, with the aim of revealing genes with
174 stage-specific over- or under-expression. We found that early and late stages had many
175 more over-expressed genes (1,434 and 1,626 respectively) than the mid stage (538)
176 (**Figure 2D, Supplementary Table S3**). A heatmap representing the expression levels
177 of the stage-specific overexpressed genes clearly separates the three groups of genes
178 (**Figure 2E**). The first group in the heatmap contains the 1,478 genes that are highly
179 expressed specifically at early stages, with less expression at mid stages and very low
180 expression at the late stage. Another large group of 1,618 genes are highly expressed
181 specifically at late stages and show low expression at early and mid stages. Finally, we
182 identified a third and smallest group of 202 genes that are highly expressed at mid stages,
183 with some detectable expression at early stages, but little detectable expression at the
184 late stage (**Figure 2E**). These results are consistent with our previous observation that
185 there is a high gene expression similarity in early and mid-stages, and an increased
186 transcriptomic change from mid to late stages.

187

188 **Separation of somatic and germ line tissues in the developing ovary**

189 Given our ultimate interest in gene regulatory functions and dynamics during terminal
190 filament formation, we wished to understand the predicted functions of the many
191 differentially expressed genes across stages. We reasoned, however, that given the
192 different developmental numbers, roles and behaviors of germ line and somatic cells in
193 this developing organ, considering functional categories of differentially expressed genes

in these whole ovary samples would be only minimally informative. We therefore designed an experimental strategy that allowed us to consider the transcriptional dynamics of the germ line and soma separately, described below.

To understand the gene expression differences between the somatic and germ line tissues of the ovary during terminal filament morphogenesis, we drove somatic and germ line tissue-specific GFP expression using the UAS-GAL4 system (BRAND AND PERRIMON 1993), using the drivers *bab:GAL4* and *nos:GAL4* respectively (see Methods). The paralogous genes *bab 1* and *bab 2* are expressed in somatic ovarian cells and are essential for terminal filament formation, differentiation of terminal filament cells and stem cell niche maintenance (SAHUT-BARNOLA *et al.* 1995b; COUDERC *et al.* 2002; SALER *et al.* 20120). The two Bab proteins act in a synergistic and partially redundant fashion in the ovary, and mutations in the *bab* genes cause defects in terminal filament and ovariole number (GODT AND LASKI 1995a; SAHUT-BARNOLA *et al.* 1995a; COUDERC *et al.* 2002). The gene *bab1* is expressed strongly in the apical and terminal filaments of the larval ovary, while *bab2* is expressed in swarm cells, in addition to other somatic cells of the larval ovary (CABRERA *et al.* 2002; COUDERC *et al.* 2002). We used the *bab pGAL$^{4\text{-}2}$* line, which was made by replacing a LacZ-carrying P-element insertion with a Gal4 element (P [Gal4,w+]) into the *bab^{P}* site in the intronic region of *bab1* gene (CABRERA *et al.* 2002). The driver *bab:GAL4* (genotype*: w[*]; P{bab1[Pgal$^{4\text{-}2}$]/TM6B, Tb[1]*) is expressed in all larval ovarian somatic cells (CABRERA *et al.* 2002; COUDERC *et al.* 2002; SARIKAYA *et al.* 2012; SALER *et al.* 2020). We used this *bab:GAL4* transgenic line to express GFP in larval ovarian somatic cells, allowing us to separate the somatic tissue from the germ line tissue using Fluorescence-Activated Cell Sorting (FACS).

The expression of the gene *nanos* (*nos*) is limited to germ line cells in the larval ovary (WANG AND LIN 2004). We used a *nos:Gal4* transgenic line (genotype: *P{w[+mC]=UAS-Dcr-2.D}1, w[1118]; P{w[+mC]=GAL4-nos.NGT}40*) to express GFP exclusively in the germ line cells, and thus isolate the germ line cells using FACS (TRACEY *et al.* 2000).

We dissociated ovaries at the three stages described above and isolated the GFP-positive cells at each stage using FACS. Cellular debris was eliminated with gate R1, non-singlets were eliminated by gate R2, and the R3 gate selected for GFP positive cells. A combination of the three gates yielded singlet GFP positive cells, minimizing the possibility of tissue contamination by undissociated cells. When similar number of ovaries were used to obtain sorted cells for somatic and germ line tissue-types, we found larger number of somatic cells as compared to germ cells as expected, indicating a successful separation of the desired tissue type (**Supplementary Figure S2**).

With this method, we obtained tissue-specific transcriptomes of somatic and germ line tissues at the same three stages of terminal filament development used to generate the whole ovary dataset. We sequenced three biological replicates for all datasets and retained replicates that had at least 10 million reads. The number of reads aligned to the genome ranged from 11.0 to 81.5 million. Greater than 94.09% of reads aligned in all datasets, with the single exception of one dataset (Mid-1) with 88.55 % of aligned reads **(Supplementary Table S1, Supplementary Figure S3**). The PCA analysis based on the counts normalized by variance-stabilization transformation (VST) shows a clear separation of somatic and germ cell libraries along the first principal component, suggesting a successful separation of cell types by FACS (**Figure 3A**). For the somatic

samples, the three biological replicates cluster closely together (**Figure 3A-B**) while the different stages are separated from each other in the second principal component. The structure of the dendrogram for the somatic samples resembles that of the whole ovary, in which early and mid-stages are closer to each other than either is to the late stage. As for the germ cell libraries, unlike the biological replicates of the early and late stages, the mid-stage replicates do not cluster together. A possible explanation is the low number of reads from the sample Mid-1 (**Supplementary Table S1, Supplementary Figure S3**).

To further assess the successful separation of somatic and germ cells, we checked the expression of well-known tissue-type-specific markers. The genes *nanos* and *vasa* are two genes known to be specifically expressed in germ cells in the ovary (SCHUPBACH AND WIESCHAUS 1986; LEHMANN AND NUSSLEIN-VOLHARD 1991). Both genes show higher expression in the germ cell libraries than in the somatic cell libraries (mean $\log_2$(Fold Change) of 8.37 for *nanos,* and 8.23 for *vasa*) (**Figure 3C**), confirming that the preparation and sequencing of the germ cell libraries successfully captured the germ cells and their RNAs, and suggesting that germ cells were not present (or present only at very low levels) in the somatic cell libraries. *bab1*, *bab2*, and *tj* are considered somatic gene markers (SAHUT-BARNOLA *et al.* 1995a; COUDERC *et al.* 2002). These three somatic markers display higher expression levels in our somatic libraries than in the germ cell libraries at each stage (mean $\log_2$(Fold Change) -0.31 for *bab1*, -1.31 for *bab2*, and -1.70 for *tj*) (**Figure 3D**). However, in four of the 18 libraries, either *bab1* or *bab2* (but not *tj*) showed higher expression levels in a specific germ cell library than in the somatic libraries. These specific cases were as follows: (1) one early stage germ cell replicate had higher *bab1* levels than one of the early somatic replicates; (2) two mid stage germ cell replicates had higher *bab1*

levels than the somatic replicates; (3) one late stage germ cell replicate had higher *bab1* levels than the somatic replicates; (4) one mid stage germ cell replicate had higher *bab2* levels than the somatic replicates. This could indicate that some somatic cells might have been included in these particular germ cell libraries. Nonetheless, despite this putative small amount of contamination, we can clearly differentiate both tissue types based on their expression profiles as shown in the PCA (**Figure 3A**), suggesting that we captured the transcriptional differences between cell types (**Figure 3A**) sufficiently to allow us to achieve our goal of successfully retrieving the genes that are highly and differentially expressed in each of these two tissues.

Differential expression analysis of somatic and germ line tissues across all stages

The differential expression analysis between the somatic and germ line tissues across all three stages revealed 1,880 genes significantly upregulated (adjusted p-value<0.01) in germ cells and 1,585 genes significantly upregulated in the somatic cells (**Figure 4A; Supplementary Table S4**).

Among the 20 most significant genes (with the lowest adjusted p-value) overexpressed in germ cells relative to somatic cells, we detected known germ line-specific genes including piRNA biogenesis genes *Argonaute3 (AGO3)*, *krimper (krimp)*, and *tejas (tej)*, along with *Aubergine (aub)*, (BRENNECKE *et al.* 2007; OLIVIERI *et al.* 2010; PATIL AND KAI 2010; SATO *et al.* 2015), *sisters unbound (sunn)* (KRISHNAN *et al.* 2014), *benign gonial cell neoplasm (bgcn)* (OHLSTEIN *et al.* 2000), and uncharacterized genes including *CG32814* and *CG12851* on the chromosome 2R. As for the somatic cells, the most significantly overexpressed gene relative to the germ cells is the cytochrome gene

Cyp4p2, whose role is unknown in the ovary, followed by cytochrome *Cyp4p1* and the uncharacterized genes CG32581 and CG42329. Some genes known to play roles in the ovary were also among this group, including the regulator of the niche cells and ecdysone receptor *Taiman (tai)* (KÖNIG *et al.* 2011), and the regulator of vitellogenesis *apterous (ap)* (GAVIN AND WILLIAMSON 1976).

Temporally dynamic expression of genes previously studied in somatic ovary development

We explored the expression dynamics of some of the previously studied genes expressed in the *Drosophila* ovary. To our knowledge, temporal gene expression studies in the larval ovary for many these genes have not yet been conducted.

First, we considered the temporal expression patterns of some adhesion proteins known to play a role in ovary development. *RanBPM* is an adhesion linker protein expressed in the germ line niche in the adult ovary (DANSEREAU AND LASKO 2008). In our dataset we see opposing trends of expression levels in somatic and germ line tissues, such that in germ line tissue *RanBPM* expression decreases progressively from early to mid to late stages, while in the somatic tissue it increases from early to mid to late stages (**Supplementary Figure S4A**). Cofilin (encoded by the gene *twinstar*) is an adhesion protein required for terminal filament cell rearrangement during terminal filament morphogenesis, as well as for adult border cell migration (CHEN *et al.* 2001). Cofilin shows similar germ line and somatic cell expression trends, with higher levels at early stages that decrease progressively at mid and late stages (**Supplementary Figure S4B**).

We then looked at temporal expression of *RhoGEF64C* and *Wnt4,* genes involved in cell motility. RhoGEF64C is a small apically localized RhoGTPase that regulates cell

shape and migration in the ovary (SIMOES *et al.* 2006). In our datasets we found *RhoGEF64C* expressed at higher levels in early and late stage somatic cells than at mid stages (**Supplementary Figure S4C**). *Wnt4* is involved in cell motility during ovarian morphogenesis (COHEN *et al.* 2002) and is expressed in the posterior terminal filaments and other somatic cell types of the third instar larval ovary (SLAIDINA *et al.* 2020). We found *Wnt4* to be expressed in lower levels in early and middle stages while the expression increases significantly in the late stage (**Supplementary Figure S4D**).

We also examined the temporal expression dynamics of a number of terminal filament cell-type-specific genes previously identified in a single cell sequencing study of the late third larval instar ovary (SLAIDINA *et al.* 2020). For example, *Diuretic hormone 44 receptor 2 (Dh442)* was identified as highly expressed in terminal filament cells (SLAIDINA *et al.* 2020). In our datasets, we observed a significant increase in expression levels only at the late stage relative to early and mid-stage expression levels (**Supplementary Figure S4E**). Additional genes known to function in terminal filaments are *engrailed, invected, hedgehog* and *patched* (FORBES *et al.* 1996; BESSE *et al.* 2005; BOLÍVAR *et al.* 2006; SALER *et al.* 2020). In our datasets we observed *engrailed* and its paralog *invected* expressed at lowest levels at the early stage, showing a progressive increase in expression levels from mid to late stages (**Supplementary Figure S4 F-G**). Interestingly, *invected,* but not *engrailed,* showed significant differential expression between early and mid-stages. The genes *patched* and *hedgehog* also showed significant increase from early to mid-stage (**Supplementary Figure S4H-I**).

Finally, we considered members of the *Fibroblast Growth Factor (FGF)* signaling pathway, which controls sheath cell proliferation in the pupal ovary (IRIZARRY AND

STATHOPOULOS 2015). Three key genes of this pathway, the FGF ligand *thisbe,* the FGF scaffolding protein *stumps* and the upstream FGF signaling activator *heartless*, show significantly higher differential expression levels at early to mid-stage than at mid to late stages (**Supplementary Figure S4J-L**). These temporal profiles add to our understanding of the roles of these genes in ovarian morphogenesis by suggesting distinct putative critical regulatory periods for different genetic pathways.

Functional enrichment analysis of differentially expressed genes in somatic and germ line tissues across all stages

To gain insight into the general functional categories of genes likely involved in ovarian germ cell and somatic behaviors during terminal filament development, we performed a gene ontology (GO) enrichment analysis of the biological processes of differentially expressed genes across cell types and developmental stages (ASHBURNER *et al.* 2000). We found 31 level four GO-terms enriched (adjusted p-value<0.05) within the upregulated genes in germ cells, and 188 level four GO-terms enriched in the upregulated genes in somatic cells (**Supplementary Figure S5**). This analysis highlighted clear differences in the biological functions performed by the genes expressed in each tissue. The GO-terms enriched in the germ cells are primarily related to meiotic processes (9/31 contain the words “meiosis” or “meiotic”), chromosome stability (6/31 contain the words “chromosome” or “karyosome”) and cell cycle (12/31 contain "cell cycle”). In contrast, the GO-terms enriched in the somatic cells are principally related to cellular response (21/188 contain “response”), development (18/188), growth (16/188), morphogenesis (10/188) cell migration (6/188 contain the word “migration”) and signaling pathways (6/188).

To complement this GO enrichment analysis, we performed a KEGG pathway enrichment analysis on the same cell-type-specific overexpressed genes. The KEGG pathway database is a manually curated database of molecular interactions used to study enrichment of genetic regulatory pathways in gene lists (KANEHISA AND GOTO 2000). With this analysis, we identified nine KEGG pathways significantly enriched in the germ cells, and 16 significantly enriched pathways in the somatic cells (adjusted p-value<0.05) (**Figure 4B**). The KEGG pathways enriched in the germ cells are generally related to meiosis and genome protection, while upregulated genes in the somatic cells are enriched for pathways involved in cell proliferation and cell death, including the previously identified *Hippo* (BARRY AND CAMARGO 2013; SARIKAYA AND EXTAVOUR 2015; ZHENG AND PAN 2019) and *MAPK* (SHAUL AND SEGER 2007) signaling pathways.

Stage- and tissue-specific differential gene expression analysis

To explore the functions of the stage-specific upregulated genes in each tissue type, we performed a differential expression analysis of somatic versus germ tissue at each of the three stages (**Supplementary Figure S6; Supplementary Table S5**) and then performed a GO analysis of biological functions and KEGG pathway enrichment analysis on the six sets of differentially expressed genes (upregulated at early, mid, and late stages in germ and somatic cells). The GO enrichment analysis of the genes differentially expressed in somatic cells over time (**Supplementary Figure S6A-C**) revealed that four key biological processes are consistent throughout all three stages including the mid stage, which has the smallest number of differentially expressed genes across stages. Specifically, these are the GO terms taxis, cell growth, actin filament-based process and

cell adhesion. At early and late stages, we additionally observe many key biological processes related to morphogenesis in the somatic cells, including cell proliferation, differentiation and migration.

To obtain a finer-grained view of the dynamic regulation of ovary development during terminal filament formation, we also performed differential expression analysis and functional enrichment analysis of somatic and germ line tissue types at each of the three stages. In the somatic cells the number of differentially expressed genes between the early and mid-stages (867 genes) is lower than between the mid and late stages (1,404 genes) (**Figure 5A; Supplementary Table S6**). To identify genes with stage-specific upregulation, we compared each stage to the other two stages. We identified a higher number of stage-specific upregulated genes in early (1,227) and late stages (1,409) than at mid (139) (**Figure 5B; Supplementary Table S7**).

The germ cells, in general, display fewer differentially expressed genes between stages than the somatic cells. From early to mid-stages there are twice as many differentially expressed genes (557 genes) as from mid to late stages (248 genes) (**Figure 5C; Supplementary Table S8**). In terms of stage-specific upregulated genes, the highest number of such genes are found at early stages (209), followed by mid (186), and late (84) stages (**Figure 5D; Supplementary Table S9**). The 1,227, 139, and 1,409 genes found upregulated at early, mid, and late stages respectively of somatic tissue were enriched for two KEGG pathways at early stages, one at mid-stage and 17 in late stages (**Figure 5E**). This analysis allowed us to pinpoint the stage(s) at which specific pathways were enriched in somatic cells relative to germ cells, which included Apoptosis, *Hippo* signaling, and *MAPK* signaling. In addition, we detected some signaling pathways

enriched only in somatic cells at late stages, such as the *Hedgehog*, *FoxO*, and *Notch* pathways (**Figure 5E**).

Given the known role of the Hippo pathway in cell proliferation (WU *et al.* 2003; HUANG *et al.* 2005; BARRY AND CAMARGO 2013), and specifically in terminal filament cell and terminal filament number regulation (SARIKAYA AND EXTAVOUR 2015), we proceeded to analyze the expression patterns of the genes belonging to the core Hippo signaling pathway. We found that most Hippo pathway core genes display increasing expression levels from early to mid to late stages, with the exception of the expression of the core gene *Rae1* which progressively decreases in expression level from early to late stages (**Supplementary Figure S7**).

In the germ cells, across stages we find fewer processes directly involved in development and morphogenesis with gene ontology categories belonging to meiosis and cell cycle (**Supplementary Figure S6**). Among the 209, 186, and 84 upregulated genes in germ cells at early, mid, and late stages respectively, only one KEGG pathway (Ribosome) and one biological process GO-term (cytoplasmic translation) were found significantly enriched at early-stages. No such enrichment was detected at mid-stages, and three KEGG pathways were enriched at late stages (**Supplementary Figure S8**).

Uncharacterized genes

The detection of uncharacterized genes among the top differentially expressed genes in germ cells drove us to ask if there were any differences in the proportion of uncharacterized genes in each set of differentially expressed genes. We found that in the genes significantly upregulated in somatic cells compared to germ cells, 29.15% are categorized as “uncharacterized proteins” in FlyBase (LARKIN *et al.* 2021), while within the

significantly upregulated genes in germ cells, the proportion of uncharacterized genes was 39.10%. Within the stage-specific upregulated genes, the proportion of uncharacterized genes remained constant (between 28.96% and 29.63%) in somatic cells, while in germ cells it increased from 29.08% in early stages, to 34.83% in mid stages, and to 37.40% in late stages (**Supplementary Figure S9).**

Expression of cell type-specific markers

A previous single cell RNA-sequencing dataset of the late third stage larval ovary (SLAIDINA *et al.* 2020) identified transcriptional profile clusters interpreted as indicative of cell types, and suggested gene markers associated with each cell type. To determine whether the cell types identified at this late stage might also be present at earlier developmental stages than that previously assessed, we examined the expression levels of those suggested marker across our datasets. As expected, the majority of the germ cell markers are highly expressed in our germ cell libraries and expressed only at low levels in the somatic cells (**Supplementary Figure S10**). Among the somatic markers detected in our somatic tissue libraries, we do not observe any particular temporal expression pattern specific to a given somatic cell type. Nevertheless, we clearly distinguish two groups of somatic markers (**Supplementary Figure S11**). One group is composed of somatic markers whose expression levels are highest at early and mid-stages, and decay at the late stage, and a larger group of makers that are less strongly expressed at early stages, show increased expression at the mid stage, and show highest expression at late stages. By contrast, the germ cell markers detected in our germ cell libraries do not display any clear temporal expression pattern. Instead, most of these

447 genes were expressed at similar levels across the three studied stages (**Supplementary**
448 **Figure S12**). This is consistent with our previous observations that the germ line dataset
449 is not enriched for any signaling pathway directly implicated in development during these
450 three times points as the somatic cells do (**Supplementary Figure S6**).

451

452 **DISCUSSION**

453 **Temporal gene expression during ovary morphogenesis**

454 We systematically staged and sequenced entire larval ovaries to generate a gene
455 expression dataset during terminal filament formation. We then separated somatic and
456 germ line tissues during these stages and generated tissue-specific transcriptomes. While
457 the development of the *Drosophila* ovary has been studied for the last several decades,
458 and progress has been made on identifying the roles of some signaling pathways in its
459 morphogenesis (COHEN *et al.* 2002; BESSE *et al.* 2005; GILBOA AND LEHMANN 2006; GANCZ
460 *et al.* 2011; GANCZ AND GILBOA 2013; MATSUOKA *et al.* 2013; GILBOA 2015; IRIZARRY AND
461 STATHOPOULOS 2015; LENGIL *et al.* 2015; MENDES AND MIRTH 2016; PANCHAL *et al.* 2017)
462 to our knowledge, there are no publicly available transcriptomes of larval ovaries of
463 *Drosophila* across developmental time. Recent articles have reported single cell RNA-
464 sequencing for *Drosophila* ovaries, focusing either on a single larval time point or on adult
465 ovaries (JEVITT *et al.* 2020; RUST *et al.* 2020; SLAIDINA *et al.* 2020; SLAIDINA *et al.* 2021).
466 Our stage and tissue-type specific data thus represent a valuable complementary
467 transcriptomic resource on the morphogenesis of the larval ovaries of *Drosophila*, a
468 complex process that ultimately influences reproductive capacity.

Differential gene expression across developmental stages of the larval ovary

Both the whole ovary datasets and the somatic tissue datasets show increased numbers of differentially expressed genes from the mid-to-late stage transition, and in the late stage of terminal filament formation in the larval ovary (**Figure 2D,5A**). In contrast, germ cells show higher numbers of differentially expressed genes in the early to-mid stage transition, and in the early stages. The similarity in differentially expressed gene numbers and signaling pathways in the whole ovary and somatic cell datasets suggests that because the somatic cells are higher in number than the germ cells (**Supplementary Figure S2**), their transcriptomes dominate the whole-ovary transcriptomes derived from late stages of larval ovary development. Further functional enrichment analyses of somatic and germ line tissue revealed that distinct functions and pathways likely operate in these two cell types during larval ovary development.

It is possible that germ cells may be especially sensitive to DNA damage given their role in propagating genetic material, which we speculate may explain the enrichment of processes related to nucleotide replication, recombination and repair in our analysis of the differentially expressed genes in germ cells **(Figures 4B, S5**). Similarly, we observed many genes of the piRNA pathway (e.g., *AGO3*, *aub*, *krimp*, *tej*), which protect the genome from transposable elements (**Supplementary Table S5**) (SATO AND SIOMI 2020) among the top significantly enriched genes in germ cells. On the other hand, the somatic tissue is enriched for different signaling pathways including Hippo, MAPK, and apoptosis (**Figures 5E, S6**), which are known to play a role in either larval or adult ovary morphogenesis (LYNCH *et al.* 2010; KHAMMARI *et al.* 2011; ELSHAER AND PIULACHS 2015; SARIKAYA AND EXTAVOUR 2015). The observation of a higher number of uncharacterized

genes in the germ line tissue datasets (**Supplementary Figure S9**) highlights the importance of future functional characterization of these genes to understand their possible roles in germ line gene regulation.

Cell adhesion and migration during ovary morphogenesis

We assessed the temporal dynamics of genes expressed in specific cell types during development to serve as generators of new hypotheses to understand the role of genes and pathways during morphogenesis. *RhoGEF64C* is a RhoGTPase with some role in regulating control cell shape changes that lead to epithelial cell invagination (SIMOES *et al.* 2006; TORET AND LE BIVIC 2021). In a genome-wide association study on ovariole number phenotypes in natural populations of *Drosophila*, *RhoGEF64C* driven in somatic tissue had a significant effect on adult ovariole number (LOBELL *et al.* 2017). The significant increase in expression of *RhoGEF64C* we observed in early and late stages (**Figure S4C**) suggests its role in somatic cell shape and migration in both early and late stages.

GO-terms related to cell adhesion, motility and taxis were enriched in all three stages in somatic cells (**Supplementary Figure S6**). Previous studies have shown signaling pathways involved in ovary development to affect cell adhesion and migration processes (COHEN *et al.* 2002; LI *et al.* 2003; BESSE *et al.* 2005; LAI *et al.* 2017). Migratory events in mid to late stages of the larval ovary have been described for two ovarian cell types, swarm cells and sheath cells (SAHUT-BARNOLA *et al.* 1995a; SAHUT-BARNOLA *et al.* 1996; GREEN II AND EXTAVOUR 2012; SLAIDINA *et al.* 2020). The increased number of differentially expressed genes that correspond to the processes of cell migration and

adhesion may be due to the migratory events in the mid- to late stages of larval ovary morphogenesis.

The FGF signaling pathway supports terminal filament cell differentiation in the early larval stages through *thisbe (ths)* and upstream activator *heartless (htl)*, and also controls sheath cell proliferation in late larval and pupal stages (IRIZARRY AND STATHOPOULOS 2015). In our dataset, we observe that in somatic cells these FGF pathway genes show a significant progressive upregulation from early to mid and from mid to late stages (**Supplementary Figure S4J-L**). Consistently, *ths* and *stumps* were identified as markers of a distinct migratory ovarian cell population, the sheath cells (SLAIDINA *et al.* 2020). The gene *stumps* is expressed in sheath cells in stages corresponding to our "late" stage in the differentiating terminal filament cells and in pupal stages (144h AEL), of migratory sheath cells (IRIZARRY AND STATHOPOULOS 2015).

Functional enrichment analysis and signaling pathways

Our results show that in the late stage of somatic cells there is an increase in expression of genes involved in multiple signaling pathways, including the Wnt, MAPK, Hippo, Hedgehog, FoxO, TGF and Notch pathways (**Figure 5E**). The molecular mechanisms of all these signaling pathways during larval ovary development have not yet been extensively studied, but all of them have been functionally implicated in ovariole number determination by a large-scale genetic screen (KUMAR *et al.* 2020).

We previously showed that Hippo signaling pathway controls proliferation of somatic cells, which affects terminal filament number (SARIKAYA AND EXTAVOUR 2015).

Our differential gene expression data show that members of the Hippo pathway are significantly differentially expressed in the somatic tissue (**Figure 4B; Supplementary Figure S8**). Loss of function mutations in Yki, an effector of the Hippo signaling pathway, cause increased growth and reduced apoptosis through an increase in the levels of the cell cycle protein Cyc E and the apoptosis inhibitor Diap1 (HARVEY *et al.* 2003; HUANG *et al.* 2005). In our somatic cell datasets, we observe *Diap1* transcript levels significantly increase from early to late stages, and those of *CycE* increase from mid to late stages (**Supplementary Figure S13A-B**). However, the Apoptosis KEGG pathway appears significantly enriched in the somatic late stage (**Figure 5E; Supplementary Figure S6**) . Furthermore, apoptosis-related genes *Dronc* and *Dark*, which form the apoptosome (**Supplementary Figure 13C, D**) (YUAN *et al.* 2011), are also significantly upregulated in the late stage, as are the caspases *Dcp-1*, *Drice*, and *Dredd* (**Supplementary Figure 13C, E-G**) (HARVEY *et al.* 2001). Thus, we observe both an upregulation of apoptosis and an upregulation of the apoptosis inhibition genes in late stage somatic cells. This could mean that genes controlling apoptosis both positively and negatively are acting to exert tight control of this process. Alternatively, our observations may reflect that each process is upregulated within different somatic cell types.

Cap cells and intermingled cells are somatic cells that interact with the germ cells for the maintenance of germ line stem cell niches (LI *et al.* 2003; SONG *et al.* 2007). The Notch signaling pathway, enriched in the late-stage somatic dataset (**Figure 5E; Supplementary Figure S6C**), is required for cap cell fate (PANCHAL *et al.* 2017; YATSENKO AND SHCHERBATA 2021). We observed an expression level increase in Notch pathway components at late stages, suggesting that the role of the Notch pathway in cap cell fate

560 determination may be particularly important at mid to late stages of larval ovary
561 development.

562 Components of the TGFβ pathway, enriched in late stage somatic cells in our
563 dataset (**Figure 5C**), are known to contribute to ovarian development. These include the
564 Bone Morphogenetic Protein (BMP) and Activin pathways of the TGFβ pathway (PANGAS
565 AND WOODRUFF 2000; GUO AND WANG 2009). The BMP ligand *decapentaplegic* (*dpp)* was
566 previously documented as expressed in all larval ovarian somatic cells and in cap cells of
567 late third instar larval ovary (XIE AND SPRADLING 1998; SATO *et al.* 2010; SALER *et al.* 2020).
568 The expression of *dpp* in the larval ovary is dependent on the expression of *bab* genes
569 (SALER *et al.* 2020). The activin pathway, controls terminal filament cell proliferation and
570 differentiation (LENGIL *et al.* 2015). We find that the activin receptor *baboon* shows a
571 significant expression level increase in the late stage somatic cells (adjusted p-value of
572 0.002602) and could indicate its role in terminal filament cell differentiation in late stage.

573

574 **Conclusions**

575 Here we provide a dataset that explores gene expression during larval ovary development
576 and morphogenesis, which is crucial to understand how the ovary is shaped in early
577 stages to develop into a functional adult organ. This work offers a dataset for the
578 developmental biology community to probe the genetic regulation of larval ovarian
579 morphogenesis.

580

581 **MATERIALS AND METHODS**

582 **Fly Stocks**

583 Flies were reared at 25°C at 60% humidity with food containing yeast and in uncrowded
584 conditions. The following two fly lines were obtained from the Bloomington *Drosophila*
585 Stock Center: *w[*]; P{bab1[Pgal$^{4-2}$]/TM6B, Tb[1]* (abbreviated herein as *bab:GAL4*; stock
586 number 6803), *P{w[+mC]=UAS-Dcr-2.D}1, w[1118]; P{w[+mC]=GAL4-nos.NGT}40*
587 (abbreviated herein as *nos:GAL4*; stock number 25751). *w[1118], P[UAS Stinger]*
588 (abbreviated herein as *UAS: Green Stinger I*, (BAROLO *et al.* 2000) used for GFP
589 expression was a gift from Dr. James Posakony (University of California, San Diego).
590 Crosses were set with 100-200 virgin UAS females and 50-100 GAL4 males in a 180 ml
591 bottle containing 50ml standard fly media one day prior to egg laying.

592

593 **Staging larvae**

594 To obtain uniformly staged larvae for the experiments, a protocol was devised to collect
595 eggs that were near-synchronously laid, from which the larvae were then collected. To
596 obtain a desired genotype, crosses were set as described above. The cross was set at
597 25°C at 60% humidity and left overnight to mate. Hourly egg collections were set up on
598 60 mm apple juice-agar plates (9 g agar, 10 g sugar, 100 ml apple juice and 300 ml water)
599 with a pea-sized spread of fresh yeast paste (baker's yeast granules made into a paste
600 in a drop of tap water). Eggs were collected hourly for eight hours. The first two collection
601 plates were discarded to remove asynchronously laid eggs that may have been retained
602 inside the females following fertilization. Staged first instar larvae were collected into vials
603 24 hours after egg collection. Larvae at 72h AEL (hours After Egg Laying) were
604 designated as early stage, at 96h AEL as middle stage and at 120h AEL as late stage of

Terminal Filament development. For a step-by-step detailed protocol see **Supplementary File 1**.

Dissection and dissociation of larval ovary

Staged larvae were collected for dissection every hour. The head of the larva was removed with forceps and the cuticle and gut were carefully pulled with one forceps while holding the fat body with another forceps. This process left just the fat bodies in the dissection dish as long as the larvae were well fed and fattened with yeast. Ovaries located in the center of the length of each fat body were then dissected free of the fat body using an insulin syringe needle (BD 328418). Ovaries dissected clear of fat body were collected in DPBS (Thermo Fisher 14190144) and batches of 20-30 ovaries in DPBS were kept on ice until dissociation. Ovaries were harvested hourly at the appropriate times, placed on ice immediately following dissection, and maintained on ice for a maximum of four hours before dissociation and subsequent FACS processing.

Dissociation of the larval ovary required two enzymatic steps. After seven hours of dissection, batches of dissected ovaries were placed in 0.25% Trypsin solution (Thermo Fisher 25200056) for ten minutes at room temperature in the cavity of a glass spot plate (Fisher Scientific 13-748B). They were then transferred to another cavity containing 2.5 % Liberase (5 g Liberase reconstituted in 2ml nuclease free water; Sigma 5401119001) and teased apart with insulin syringe needles until most of the clumps were separated and left (without agitation) at room temperature for ten minutes. Using a 200µl pipette with a filter tip (pre rinsed in 1X PBS), the dissociated cells in Liberase were pipetted up and down gently ten times to uniformly mix and separate the cells. The cell suspension was

then transferred to an RNA Lobind tube (Eppendorf 8077-230) and placed on a vortexer for 1 minute. Meanwhile the well was rinsed in 1.4 ml of PBS by pipetting repeatedly. This PBS was then mixed with the cell-suspension in Liberase and vortexed for another minute, and the entire sample was then placed on ice. This sample was then taken directly to the FACS facility on ice along with an RNA Lobind collection tube containing 100-200µl Trizol (Thermo Fisher 15596026). For a step-by-step detailed protocol see **Supplementary File 1**.

Flow Sorting GFP-positive cells

The dissociated tissue sample was sorted in a MoFlo Astrios EQ Cell sorter (Beckman Coulter) run with Summit v6.3.1 software. The dissociated cell solution was diluted and a flow rate of 200 events per second was maintained with high sorting efficiency (< 98%) during the sorting process. A scatter gate (R1) was employed to eliminate debris (**Supplementary Figure S2**) and a doublet gate (R2) was used to exclude non-singlet cells. A 488 nm emission Laser was used to excite the GFP, and the collection was at 576 nm. The GFP-positive cells were designated in gate R3 and sorted directly into Trizol. The resulting cells collected in Trizol were frozen immediately by plunging the tube in liquid nitrogen and then stored at -80°C until RNA extraction. A single replicate consisted of at least 1000 cell counts pooled from FACS runs.

RNA extraction

649 Flow-sorted cells were stored at -80°C were thawed at room temperature. Trizol contents
650 were lysed with a motorized pellet pestle (Kimble 749540-0000). Zymo RNA Micro-Prep
651 kit (Zymo Research R2060) was used to isolate RNA from the Trizol preparations. Equal
652 amounts of molecular grade ethanol (Sigma E7023) were added to Trizol and mixed well
653 with a pellet pestle, then pipetted onto a spin column. All centrifugation steps were done
654 at 10,000g for one minute at room temperature. The column was washed with 400µl Zymo
655 RNA wash buffer and then treated with Zymo DNase (6U/µl) for 15 minutes at room
656 temperature. The column was then washed twice with 400µl Zymo RNA Pre-wash buffer
657 and once with Zymo RNA wash-buffer. The RNA was eluted from the column in 55 µl of
658 Nuclease-free water (Thermo Fisher 10977015). The RNA obtained was quantified first
659 using a NanoDrop (Model ND1000) spectrophotometer and then using a high sensitivity
660 kit (Thermo Fisher Q32852) on a Qubit 3.0 Fluorometer (Thermo Fisher Q33216). It was
661 also checked for integrity on a high sensitivity tape (Agilent 5067-5579) with an electronic
662 ladder on an Agilent Tapestation 2200 or 4200. RNA extraction from staged whole ovaries
663 was carried out by crushing entire ovaries in Trizol and following the same protocol
664 described above. For a step-by-step detailed protocol see **Supplementary File 1**.

665

666 **Library Preparation**

667 cDNA libraries were prepared using the Takara Apollo library preparation kit (catalogue
668 # 640096). Extracted RNA samples were checked for quality using Tapestation tapes.
669 50µl of RNA samples were pipetted into Axygen PCR 8-strip tubes (Fisher Scientific 14-
670 222-252) and processed through PrepX protocols on the Apollo liquid handling system.
671 mRNA was isolated using PrepX PolyA-8 protocol (Takara 640098). The mRNA samples

were then processed for cDNA preparation using PrepX mRNA-8 (Takara 640096) protocol. cDNA products were then amplified for 15 cycles of PCR using longAmp Taq (NEB M0287S). During amplification PrepX RNAseq index barcode primers were added for each library to enable multiplexing. The amplified library was then cleaned up using PrepX PCR cleanup-8 protocol with magnetic beads (Aline C-1003). The final cDNA libraries were quantified using a high sensitivity dsDNA kit (Thermo Fisher Q32854) on a Qubit 3.0 Fluorometer (Thermo Fisher Q33216). cDNA content and quality were assessed with D1000 (Agilent 5067-5582) or High sensitivity D1000 tape (Agilent 5067-5584, when cDNA was in low amounts) on an Agilent Tapestation 2200 or 4200. For a step-by-step detailed protocol see **Supplementary File 1**.

Sequencing cDNA libraries

Libraries were sequenced on an Illumina HiSeq 2500 sequencer. Single end-50bp reads were sequenced on a high-throughput flow cell. Libraries of varying concentrations were normalized to be equimolar, the concentrations of which ranged between 2-10nM per lane. All the samples in a flow cell were multiplexed and later separated based on unique prepX indices to yield at least 10 million reads per library. The reads were demultiplexed and trimmed of adapters using the bcl2fastq2 v2.2 pipeline to yield final fastq data files.

RNA-seq data processing

The *D. melanogaster* genome assembly and gene annotations were obtained from FlyBase version dmel_r6.36_FB2020_05 (LARKIN *et al.* 2021). The reads were aligned

with RSEM v1.3.3 (LI AND DEWEY 2011) and using STAR v2.7.6a as read aligner (DOBIN *et al.* 2013) we obtained the gene counts in each library. Because some of the tissue-specific biological samples were sequenced in more than one lane or run, and therefore the reads were split into multiple fastq files, the gene counts belonging to the same biological sample were summed. Gene counts in each dataset were normalized with the variance stabilizing transformation (VST) method implemented in the DESeq2 v1.26.0 (LOVE *et al.* 2014) R package. Further analyses, such as principal component analysis, hierarchical clustering, and differential expression analysis, were performed in R using the VST-normalized counts.

Differential Expression (DE) analysis

The differential expression analyses were performed with DESeq2 v1.26.0 (LOVE *et al.* 2014). On the whole ovary dataset, the contrasts tested were early vs mid, and mid vs late stages. For the tissue-specific datasets, three different comparisons were performed. First, to identify differentially expressed genes independently of the stage, all stages of somatic cells were compared to all stages of germ cells. Second, to identify genes up-regulated in a stage-specific manner within each tissue, we compared the expression level at each stage to the mean expression level of the other two stages. Third, we compared germ cells and somatic cells independently at each stage. Genes with a Benjamini-Hochberg (BH) adjusted p-value lower than 0.01 were selected as differentially expressed in the corresponding contrast.

Functional analysis

The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analyses were performed on the differentially expressed genes with the enrichGO and enrichKEGG functions of the clusterProfiler package (v3.14.3) for R (YU *et al.* 2012). The GO terms were obtained using the R package AnnotationDbi (CARLSON 2015) with the database org.Dm.eg.db v3.10.0. The GO overrepresentation analysis of biological process (BP) was performed against the gene universe of all *D. melanogaster* annotated genes in org.Dm.eg.db, adjusting the p-values with the Benjamini-Hochberg method (BH), adjusted p-value and q-value cutoff of 0.01, and a minimum of 30 genes per term. For the KEGG enrichment analysis, p-values were adjusted by the BH procedure, and an adjusted p-value cutoff of 0.05 was used.

Data availability

All the raw data are publicly available at NCBI-Gene Expression Omnibus (GEO) database under the accession code GSE172015. The scripts used to process and analyze the data are available at GitHub repository https://github.com/guillemylla/Ovariole_morphogenesis_RNAseq.

Competing Interests

The authors have no competing interests to declare.

Acknowledgements

This study was supported by National Institutes of Health NICHD award R01-HD073499-01 to CGE, funds from Harvard University, and the Google Cloud Academic Research Credits Program. We thank the flow cytometry and genomics staff at the Bauer Core Facility of Harvard University Faculty of Arts and Sciences (FAS) for advice on FACS sorting and sequencing, the FAS Research Computing group for advice on initial data analysis, and Extavour lab members for discussion.

FIGURE LEGENDS

Figure 1: **Experimental scheme for generating stage-specific transcriptomes of germ cells and somatic cells of larval ovaries during terminal filament formation.** **A)** Location of the larval ovaries (white circles within the larva), and illustration of larval ovary development divided into three stages during terminal filament formation (colored in black). **B)** Left to right: location of the ovaries in an adult female abdomen; a single adult ovary containing multiple ovarioles; an individual ovariole; anterior tip of an ovariole enlarged to show the germarium and terminal filament (black) at the tip. **C)** Representation of the three stages of whole larval ovaries from the wild type strain Oregon R, chosen for library preparation and sequencing (light gray: early stage, gray: mid, dark gray: late). **D)** Somatic cells and **E)** germ cells from developing ovaries at the three chosen stages were labelled with GFP using tissue specific GAL4 lines (somatic cells in shades of cyan and germ cells in shades of magenta). Somatic cells were labelled using *bab:GAL4* (genotype: *w[*]; P{w[+mW.hs]=GawB}bab1[Pgal4-2]/TM6B, Tb[1])* and germ cells were labeled using *nos:GAL4* (genotype: *P{w[+mC]=UAS-Dcr-2.D}1, w[1118]; P{w[+mC]=GAL4-nos.NGT}40)* and **F)** GFP-positive cells were separated using FACS. **G)** Schematics of representative plot layouts of somatic and germ line tissue separation using FACS. Y axis: autofluorescence, 488-576/21 Height Log; X axis: GFP fluorescence intensity, 488-513/26 Height Log (see Supplementary Figure S2 for actual representative data plots). **H)** Separated cells or whole ovaries were processed for mRNA extraction and cDNA library preparation followed by high throughput sequencing. h AEL = hours After Egg Laying.

Figure 2: Whole ovary RNA-seq dataset overview. **A)** hierarchical clustering dendrogram and **B)** PCA of the whole ovary RNA-seq dataset, both showing that biological replicates are similar to each other, and that early and mid-stages are more similar to each other than either of them is to late stage. **C)** Number of differentially expressed genes between early and mid stages, and between mid and late stages (adjusted p-value<0.01; black: upregulated genes; white: downregulated genes). See Supplementary Table S2 for gene list. **D)** Number of significantly upregulated stage-specific genes (adjusted p-value<0.01). See Supplementary Table S3 for gene list. **E)** Heatmap showing the expression of all the stage-specific upregulated genes as a row-wise z-score. Genes are clustered hierarchically and separated into three groups using the function “cutree”, and grayscale row labels (“*Gene DE”) immediately to the right of the tree are colored based on the stage in which the gene was detected to be significantly upregulated (x axis categories).

Figure 3: Cell type-specific RNA-seq dataset concordance and positive controls. **A)** PCA Plot and **B)** hierarchical clustering dendrogram of germ cell and somatic cell RNA-seq libraries. Expression in normalized counts by variance stabilization transformation (VST) in each of the cell-type-specific RNA-seq libraries of **C)** known germ cell markers *nanos* and *vasa,* and **D)** known somatic cell markers *bric a brac 1*, *bric a brac 2*, and *traffic jam*.

Figure 4: Transcriptomic differences between germ cells and somatic cells. **A)** Number of significantly upregulated genes (adjusted p-value<0.01) in germ cells and somatic cells. See Supplementary Table S4 for gene list. **B)** Significantly enriched KEGG pathways (adjusted p-value<0.05) within the upregulated genes of each cell type. The circle size is proportional to the number of differentially expressed genes that the indicated KEGG pathway contains, and the color gradient indicates the p-value.

Figure 5: Cell type-specific differential expression analysis. **A)** Number of differentially expressed genes (adjusted p-value<0.01) upregulated (black) and downregulated (white) in somatic cells between early and mid, and between mid and late stages. See Supplementary Table S6 for gene list. **B)** Number of differentially expressed genes (adjusted p-value<0.01) upregulated (black) and downregulated (white) in somatic cells at each stage compared to the two other stages. See Supplementary Table S7 for gene list. **C)** Number of differentially expressed genes (adjusted p-value<0.01) upregulated (black) and downregulated (white) in germ cells between early and mid, and between mid and late stages. See Supplementary Table S8 for gene list. **D)** Number of differentially expressed genes (adjusted p-value<0.01) upregulated (black) and downregulated (white) in germ cells at each stage compared to the two other stages. See Supplementary Table S9 for gene list. **E)** Significantly enriched (adjusted p-value<0.05) KEGG pathways within the upregulated genes at each somatic stage. Circle size is proportional to the number of differentially expressed genes it contains, and the color gradient indicates the p-value.

REFERENCES

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics 25**:** 25-29.

Ashburner, M., K. G. Golic and R. S. Hawley, 2005 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Barolo, S., L. A. Carver and J. W. Posakony, 2000 GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. Biotechniques 29**:** 726, 728, 730, 732.

Barry, E. R., and F. D. Camargo, 2013 The Hippo superhighway: signaling crossroads converging on the Hippo/Yap pathway in stem cells and development. Current Opinion in Cell Biology 25**:** 247-253.

Besse, F., D. Busson and A. M. Pret, 2005 Hedgehog signaling controls Soma-Germen interactions during Drosophila ovarian morphogenesis. Dev Dyn 234**:** 422-431.

Bolivar, J., J. Pearson, L. Lopez-Onieva and A. Gonzalez-Reyes, 2006 Genetic dissection of a stem cell niche: the case of the Drosophila ovary. Dev Dyn 235**:** 2969-2979.

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

Brand, A. H., and N. Perrimon, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118**:** 401-415.

Brennecke, J., A. A. Aravin, A. Stark, M. Dus, M. Kellis *et al.*, 2007 Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell 128**:** 1089-1103.

Büning, J., 1994 *The Insect Ovary: ultrastructure, previtellogenic growth and evolution*. Chapman and Hall, London.

Cabrera, G. R., D. Godt, P. Y. Fang, J. L. Couderc and F. A. Laski, 2002 Expression pattern of Gal4 enhancer trap insertions into the bric a brac locus generated by P element replacement. Genesis 34**:** 62-65.

Carlson, M., 2015 AnnotationDbi: Introduction To Bioconductor Annotation Packages.

Chen, J., D. Godt, K. Gunsalus, I. Kiss, M. Goldberg *et al.*, 2001 Cofilin/ADF is required for cell motility during Drosophila ovary development and oogenesis. Nat Cell Biol 3**:** 204-209.

Church, S. H., B. A. S. de Medeiros, S. Donoughe, N. L. Marquez Reyes and C. G. Extavour, 2021 Repeated loss of variation in insect ovary morphology highlights the role of developmental constraint in life-history evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences in press.

Cohen, E. D., M. C. Mariol, R. M. Wallace, J. Weyers, Y. G. Kamberov *et al.*, 2002 DWnt4 regulates cell movement and focal adhesion kinase during Drosophila ovarian morphogenesis. Dev Cell 2**:** 437-448.

Couderc, J. L., D. Godt, S. Zollman, J. Chen, M. Li *et al.*, 2002 The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development 129**:** 2419-2433.

Dansereau, D. A., and P. Lasko, 2008 RanBPM regulates cell shape, arrangement, and capacity of the female germline stem cell niche in Drosophila melanogaster. J Cell Biol 182**:** 963-977.

David, J., 1970 Le nombre d'ovarioles chez la Drosophila: relation avec la fecondite et valeur adaptive [Number of ovarioles inDrosophila melanogaster: relationship with fertility and adaptive value]. Arch. Zool. Exp. Gen 111**:** 357-370.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner., pp. 15-21 in *Bioinformatics*.

Elshaer, N., and M. D. Piulachs, 2015 Crosstalk of EGFR signalling with Notch and Hippo pathways to regulate cell specification, migration and proliferation in cockroach panoistic ovaries. Biol Cell 107**:** 273-285.

Forbes, A. J., H. Lin, P. W. Ingham and A. C. Spradling, 1996 hedgehog is required for the proliferation and specification of ovarian somatic cells prior to egg chamber formation in Drosophila. Development 122**:** 1125-1135.

Gancz, D., and L. Gilboa, 2013 Insulin and Target of rapamycin signaling orchestrate the development of ovarian niche-stem cell units in Drosophila. Development 140**:** 4145-4154.

Gancz, D., T. Lengil and L. Gilboa, 2011 Coordinated regulation of niche and stem cell precursors by hormonal signaling. PLoS Biology 9**:** e1001202.

Gavin, J. A., and J. H. Williamson, 1976 Juvenile hormone-induced vitellogenesis in Apterous4, a non-vitellogenic mutant in Drosophila melanogaster. Journal of Insect Physiology 22**:** 1737-1742.

Gilboa, L., 2015 Organizing stem cell units in the *Drosophila* ovary. Current Opinion in Genetics & Development 32C**:** 31-36.

Gilboa, L., and R. Lehmann, 2004 Repression of primordial germ cell differentiation parallels germ line stem cell maintenance. Curr Biol 14**:** 981-986.

Gilboa, L., and R. Lehmann, 2006 Soma-germline interactions coordinate homeostasis and growth in the Drosophila gonad. Nature 443**:** 97-100.

Godt, D., and F. A. Laski, 1995a Mechanisms of cell rearrangement and cell recruitment in Drosophila ovary morphogenesis and the requirement of bric a brac. Development 121**:** 173-187.

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

Green, D. A., 2nd, and C. G. Extavour, 2014 Insulin signalling underlies both plasticity and divergence of a reproductive trait in Drosophila. Proc Biol Sci 281**:** 20132673.

Green II, D. A., and C. G. Extavour, 2012 Convergent Evolution of a Reproductive Trait Through Distinct Developmental Mechanisms in *Drosophila*. Developmental Biology 372**:** 120-130.

Guo, X., and X. F. Wang, 2009 Signaling cross-talk between TGF-beta/BMP and other pathways. Cell Res 19**:** 71-88.

Harvey, K. F., C. M. Pfleger and I. K. Hariharan, 2003 The Drosophila Mst ortholog, hippo, restricts growth and cell proliferation and promotes apoptosis. Cell 114**:** 457-467.

Harvey, N. L., T. Daish, K. Mills, L. Dorstyn, L. M. Quinn *et al.*, 2001 Characterization of the Drosophila caspase, DAMM. J Biol Chem 276**:** 25342-25350.

Hodin, J., 2009 Chapter 11: She Shapes Events As They Come: Plasticity in Female Insect Reproduction, pp. 423-521 in *Phenoypic Plasticity of Insects*, edited by D. Whitman and T. N. Ananthakrishnan. Science Publishers, Inc., Enfield, NH.

Hodin, J., and L. M. Riddiford, 2000a Different mechanisms underlie phenotypic plasticity and interspecific variation for a reproductive character in drosophilids (Insecta: Diptera). Evolution 54**:** 1638-1653.

Hodin, J., and L. M. Riddiford, 2000b Different mechanisms underlie phenotypic plasticity and interspecific variation for a reproductive character in Drosophilids (Insecta: Diptera). Evolution 5**:** 1638-1653.

Honek, A., 1993 Intraspecific Variation in Body Size and Fecundity in Insects - a General Relationship. Oikos 66**:** 483-492.

Huang, J., S. Wu, J. Barrera, K. Matthews and D. Pan, 2005 The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila Homolog of YAP. Cell 122**:** 421-434.

Irizarry, J., and A. Stathopoulos, 2015 FGF signaling supports Drosophila fertility by regulating development of ovarian muscle tissues. Dev Biol 404**:** 1-13.

Jevitt, A., D. Chatterjee, G. Xie, X. F. Wang, T. Otwell *et al.*, 2020 A single-cell atlas of adult Drosophila ovary identifies transcriptional programs and somatic cell lineage regulating oogenesis. PLoS Biol 18**:** e3000538.

Kanehisa, M., and S. Goto, 2000 KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28**:** 27-30.

Khammari, A., F. Agnes, P. Gandille and A. M. Pret, 2011 Physiological apoptosis of polar cells during Drosophila oogenesis is mediated by Hid-dependent regulation of Diap1. Cell Death Differ 18**:** 793-805.

King, R. C., 1970 *Ovarian development in Drosophila melanogaster*. Academic Press.

King, R. C., S. K. Aggarwal and U. Aggarwal, 1968 The Development of the Female *Drosophila* Reproductive System. Journal of Morphology 124**:** 143-166.

König, A., A. S. Yatsenko, M. Weiss and H. R. Shcherbata, 2011 Ecdysteroids affectDrosophilaovarian stem cell niche formation and early germline differentiation. The EMBO Journal 30**:** 1549-1562.

Krishnan, B., S. E. Thomas, R. Yan, H. Yamada, I. B. Zhulin *et al.*, 2014 Sisters Unbound Is Required for Meiotic Centromeric Cohesion in Drosophila melanogaster. Genetics 198**:** 947-965.

Kumar, T., L. Blondel and C. G. Extavour, 2020 Topology-driven analysis of protein-protein interaction networks detects functional genetic sub-networks regulating reproductive capacity. eLife 9**:** e54082.

Lai, C. M., K. Y. Lin, S. H. Kao, Y. N. Chen, F. Huang *et al.*, 2017 Hedgehog signaling establishes precursors for germline stem cell niches by regulating cell adhesion. Journal of Cell Biology 216**:** 1439-1453.

Larkin, A., S. J. Marygold, G. Antonazzo, H. Attrill, G. Dos Santos *et al.*, 2021 FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res 49**:** D899-D907.

Leatherman, J. L., and T. A. Jongens, 2003 Transcriptional silencing and translational control: key features of early germline development. Bioessays 25**:** 326-335.

Lehmann, R., and C. Nusslein-Volhard, 1991 The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. Development 112**:** 679-691.

Lengil, T., D. Gancz and L. Gilboa, 2015 Activin signaling balances proliferation and differentiation of ovarian niche precursors and enables adjustment of niche numbers. Development 142**:** 883-892.

Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome., pp. 323 in *BMC Bioinformatics*.

Li, M., X. Hu, S. Zhang, M. S. Ho, G. Wu *et al.*, 2019 Traffic jam regulates the function of the ovarian germline stem cell progeny differentiation niche during pre-adult stage in Drosophila. Sci Rep 9**:** 10124.

Li, M. A., J. D. Alls, R. M. Avancini, K. Koo and D. Godt, 2003 The large Maf factor Traffic Jam controls gonad morphogenesis in *Drosophila*. Nature Cell Biology 5**:** 994-1000.

Lobell, A. S., R. R. Kaspari, Y. L. Serrano Negron and S. T. Harbison, 2017 The Genetic Architecture of Ovariole Number in Drosophila melanogaster: Genes with Major, Quantitative, and Pleiotropic Effects. G3 (Bethesda) 7**:** 2391-2403.

Love, M. I., W. Huber and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15**:** 550.

Lynch, J. A., A. D. Peel, A. Drechsler, M. Averof and S. Roth, 2010 EGF signaling and the origin of axial polarity among the insects. Curr Biol 20**:** 1042-1047.

Markow, T. A., and P. M. O'Grady, 2007 *Drosophila* biology in the genomic age. Genetics 177**:** 1269-1276.

Matsuoka, S., Y. Hiromi and M. Asaoka, 2013 Egfr signaling controls the size of the stem cell precursor pool in the Drosophila ovary. Mech Dev 130**:** 241-253.

Mendes, C. C., and C. K. Mirth, 2016 Stage-Specific Plasticity in Ovary Size Is Regulated by Insulin/Insulin-Like Growth Factor and Ecdysone Signaling in Drosophila. Genetics 202**:** 703-719.

Ohlstein, B., C. A. Lavoie, O. Vef, E. Gateff and D. M. McKearin, 2000 The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics 155**:** 1809-1819.

Olivieri, D., M. M. Sykora, R. Sachidanandam, K. Mechtler and J. Brennecke, 2010 An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J 29**:** 3301-3317.

Panchal, T., X. Chen, E. Alchits, Y. Oh, J. Poon *et al.*, 2017 Specification and spatial arrangement of cells in the germline stem cell niche of the Drosophila ovary depend on the Maf transcription factor Traffic jam. PLoS Genet 13**:** e1006790.

Pangas, S. A., and T. K. Woodruff, 2000 Activin signal transduction pathways. Trends Endocrinol Metab 11**:** 309-314.

Patil, V. S., and T. Kai, 2010 Repression of Retroelements in Drosophila Germline via piRNA Pathway by the Tudor Domain Protein Tejas. Current Biology 20**:** 724-730.

Rust, K., L. E. Byrnes, K. S. Yu, J. S. Park, J. B. Sneddon *et al.*, 2020 A single-cell atlas and lineage analysis of the adult Drosophila ovary. Nat Commun 11**:** 5628.

Sahut-Barnola, I., B. Dastugue and J.-L. Couderc, 1996 Terminal filament cell organization in the larval ovary of *Drosophila melanogaster*: ultrastructural

observations and pattern of divisions. Roux's Archives of Developmental Biology 205**:** 356-363.

Sahut-Barnola, I., D. Godt, F. A. Laski and J.-L. Couderc, 1995a *Drosophila* Ovary Morphogenesis: Analysis of Terminal Filament Formation and Identification of a Gene Required for This Process. Developmental Biology 170**:** 127-135.

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

Saler, L. M., M., V. Hauser, M. Bartoletti, C. Mallart, M. Malartre *et al.*, 2020 The Bric-à-Brac BTB/POZ transcription factors are necessary in niche cells for germline stem cells establishment and homeostasis through control of BMP/DPP signaling in the *Drosophila melanogaster* ovary. PLOS Genetics 16: e1009128.

Sarikaya, D. P., A. A. Belay, A. Ahuja, A. Dorta, D. A. Green, 2nd *et al.*, 2012 The roles of cell size and cell number in determining ovariole number in Drosophila. Dev Biol 363**:** 279-289.

Sarikaya, D. P., and C. G. Extavour, 2015 The Hippo pathway regulates homeostatic growth of stem cell niche precursors in the *Drosophila* ovary. PLoS Genetics 11**:** e1004962.

Sato, K., and M. C. Siomi, 2020 The piRNA pathway in Drosophila ovarian germ and somatic cells. Proc Jpn Acad Ser B Phys Biol Sci 96**:** 32-42.

Sato, K., Yuka, A. Shibuya, P. Carninci, Y. Tsuchizawa *et al.*, 2015 Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Molecular Cell 59**:** 553-563.

Sato, T., J. Ogata and Y. Niki, 2010 BMP and Hh signaling affects primordial germ cell division in Drosophila. Zoolog Sci 27**:** 804-810.

Schupbach, T., and E. Wieschaus, 1986 Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. Rouxs Arch Dev Biol 195**:** 302-317.

Shaul, Y. D., and R. Seger, 2007 The MEK/ERK cascade: from signaling specificity to diverse functions. Biochimica et Biophysica Acta 1773**:** 1213-1226.

Simoes, S., B. Denholm, D. Azevedo, S. Sotillos, P. Martin *et al.*, 2006 Compartmentalisation of Rho regulators directs cell invagination during tissue morphogenesis. Development 133**:** 4257-4267.

Slaidina, M., T. U. Banisch, S. Gupta and R. Lehmann, 2020 A single-cell atlas of the developing Drosophila ovary identifies follicle stem cell progenitors. Genes Dev 34**:** 239-249.

Slaidina, M., S. Gupta, T. Banisch and R. Lehmann, 2021 A single cell atlas reveals unanticipated cell type complexity in Drosophila ovaries. bioRxiv**:** 427703.

Song, X., G. B. Call, D. Kirilly and T. Xie, 2007 Notch signaling controls germline stem cell niche formation in the Drosophila ovary. Development 134**:** 1071-1080.

Toret, C. P., and A. Le Bivic, 2021 A potential Rho GEF and Rac GAP for coupled Rac and Rho cycles during mesenchymal-to-epithelial-like transitions. Small GTPases 12**:** 13-19.

Tracey, W. D., Jr., X. Ning, M. Klingler, S. G. Kramer and J. P. Gergen, 2000 Quantitative analysis of gene function in the Drosophila embryo. Genetics 154**:** 273-284.

Twombly, V., R. K. Blackman, H. Jin, J. M. Graff, R. W. Padgett *et al.*, 1996 The TGF-beta signaling pathway is essential for Drosophila oogenesis. Development 122**:** 1555-1565.

Wang, Z., and H. Lin, 2004 Nanos maintains germline stem cell self-renewal by preventing differentiation. Science 303**:** 2016-2019.

Wu, S., J. Huang, J. Dong and D. Pan, 2003 hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts. Cell 114**:** 445-456.

Xie, T., and A. C. Spradling, 1998 decapentaplegic is essential for the maintenance and division of germline stem cells in the Drosophila ovary. Cell 94**:** 251-260.

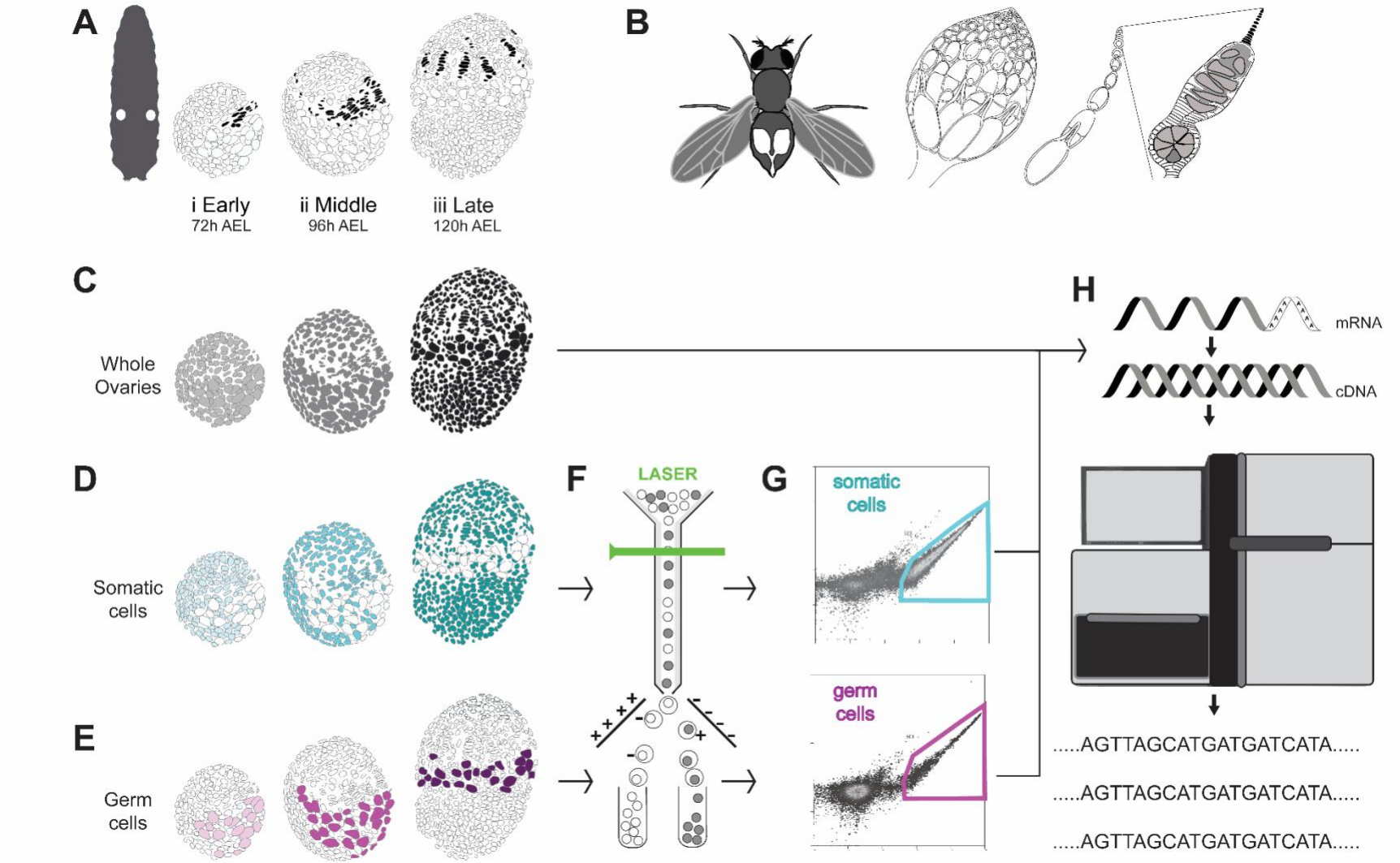
Yatsenko, A. S., and H. R. Shcherbata, 2021 Distant activation of Notch signaling induces stem cell niche assembly. PLoS Genet 17**:** e1009489.

Yu, G., L. G. Wang, Y. Han and Q. Y. He, 2012 clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16**:** 284-287.

Yuan, S., X. Yu, M. Topf, L. Dorstyn, S. Kumar *et al.*, 2011 Structure of the Drosophila apoptosome at 6.9 a resolution. Structure 19**:** 128-140.

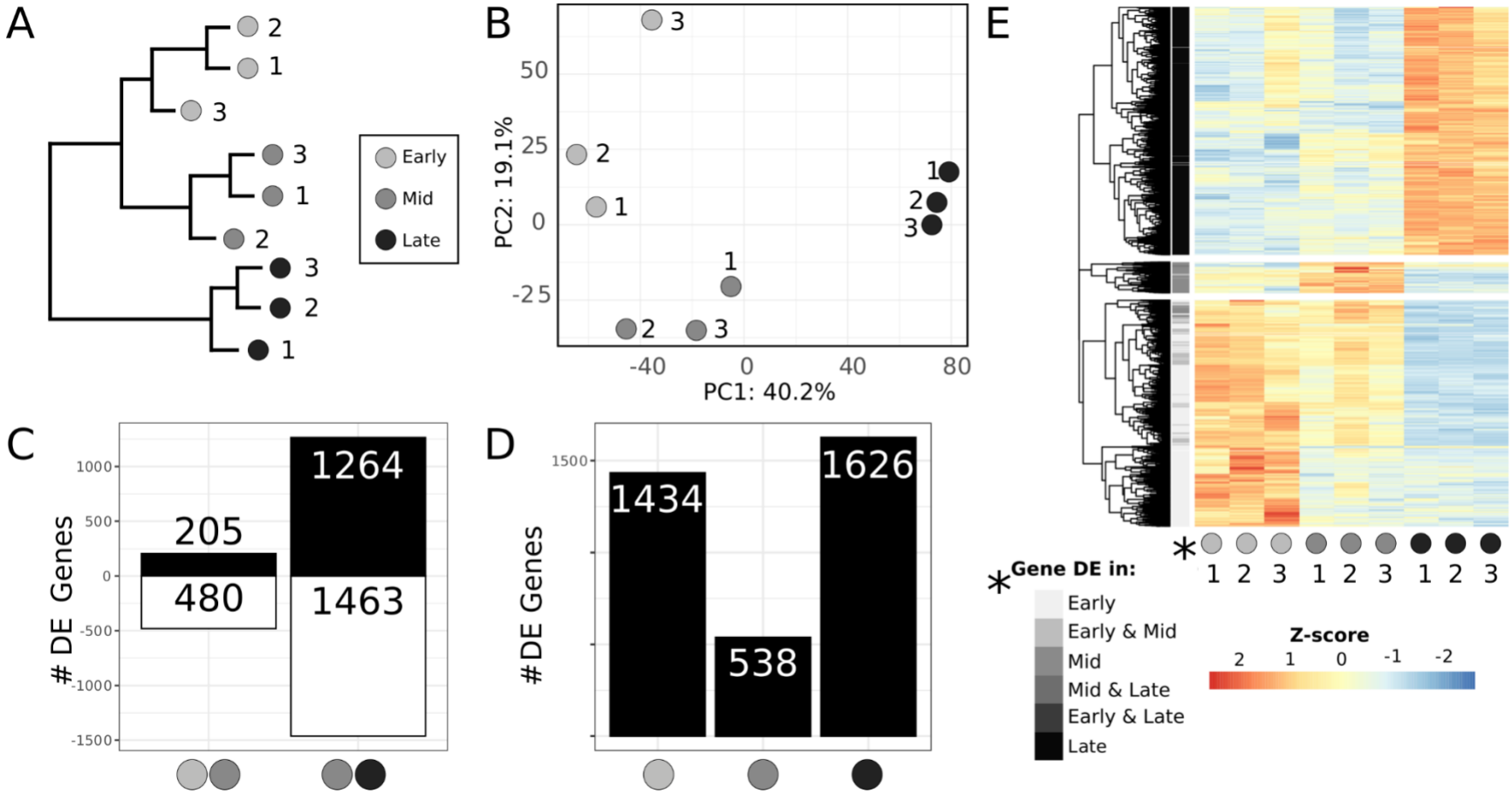
Zheng, Y., and D. Pan, 2019 The Hippo Signaling Pathway in Development and Disease. Dev Cell 50**:** 264-282.

1052 **Figure 1**



1053

Figure 2


A
2
1
3
3
1
2
3
2
1
Early
Mid
Late
B
PC2: 19.1%
PC1: 40.2%
50
25
0
-25
-40
0
40
80
C
DE Genes
1000
500
0
-500
-1000
-1500
205
480
1264
1463
D
#DE Genes
1500
1434
538
1626
E
Gene DE in:
Early
Early & Mid
Mid
Mid & Late
Early & Late
Late
Z-score
2
1
0
-1
-2
1 2 3 1 2 3 1 2 3

1057 **Figure 3**

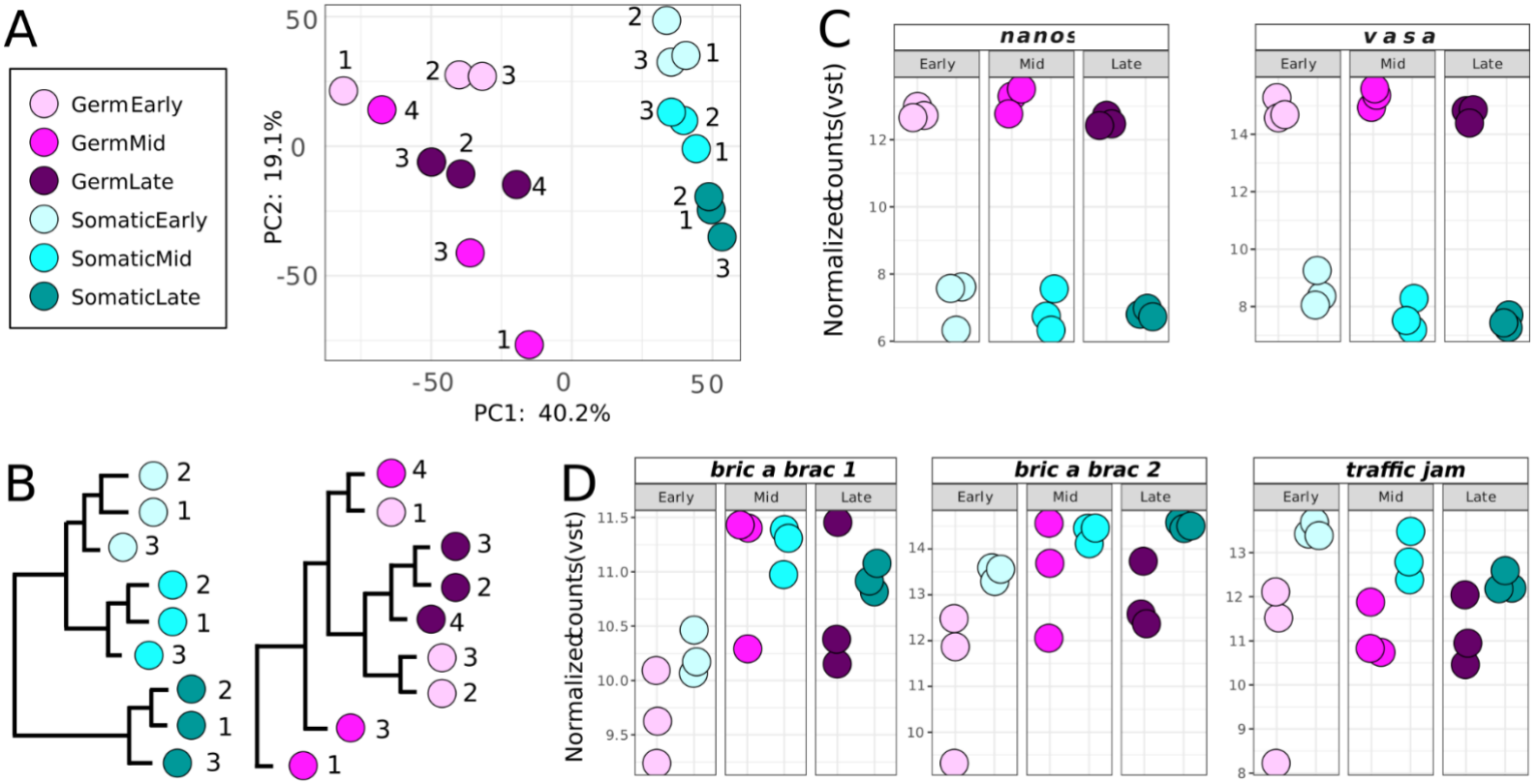


Figure 4

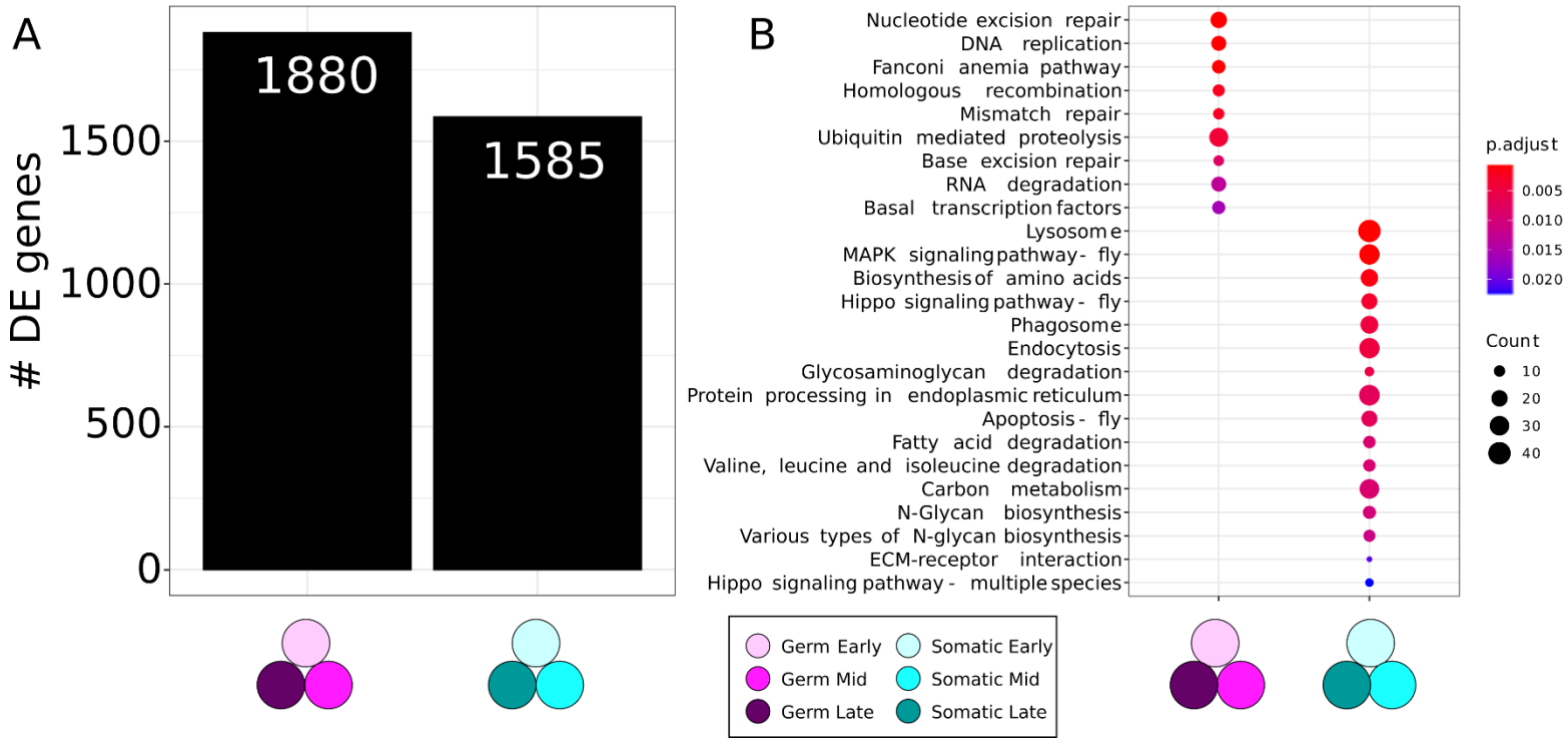
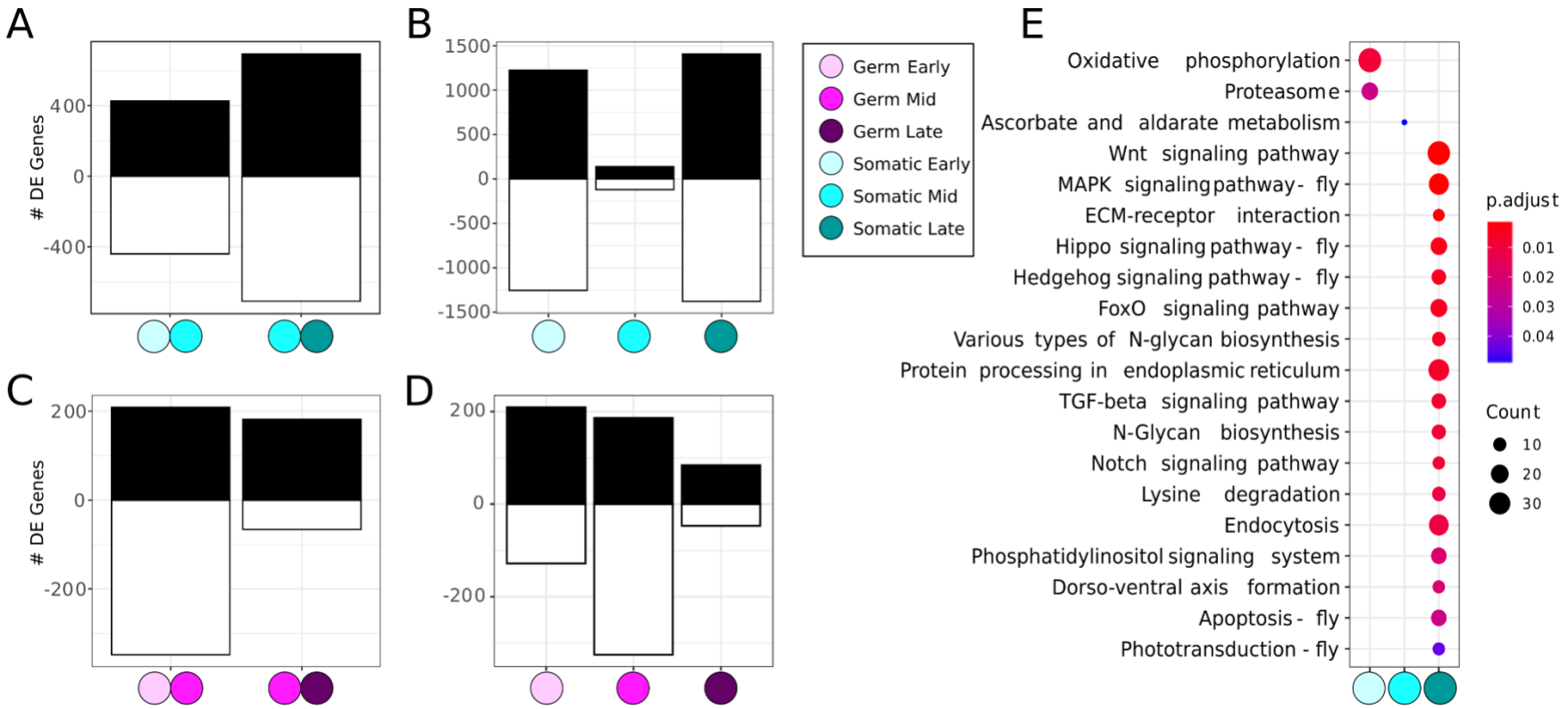

A
DE genes
1500
1000
500
0
1880
1585
B
Nucleotide excision repair
DNA replication
Fanconi anemia pathway
Homologous recombination
Mismatch repair
Ubiquitin mediated proteolysis
Base excision repair
RNA degradation
Basal transcription factors
Lysosome
MAPK signaling pathway - fly
Biosynthesis of amino acids
Hippo signaling pathway - fly
Phagosome
Endocytosis
Glycosaminoglycan degradation
Protein processing in endoplasmic reticulum
Apoptosis - fly
Fatty acid degradation
Valine, leucine and isoleucine degradation
Carbon metabolism
N-Glycan biosynthesis
Various types of N-glycan biosynthesis
ECM-receptor interaction
Hippo signaling pathway - multiple species
p.adjust
0.005
0.010
0.015
0.020
Count
10
20
30
40
Germ Early
Germ Mid
Germ Late
Somatic Early
Somatic Mid
Somatic Late

Figure 5


A
B
C
D
E
DE Genes
400
0
-400
1500
1000
500
-500
-1000
-1500
200
-200
Germ Early
Germ Mid
Germ Late
Somatic Early
Somatic Mid
Somatic Late
Oxidative phosphorylation
Proteasome
Ascorbate and aldarate metabolism
Wnt signaling pathway
MAPK signaling pathway - fly
ECM-receptor interaction
Hippo signaling pathway - fly
Hedgehog signaling pathway - fly
FoxO signaling pathway
Various types of N-glycan biosynthesis
Protein processing in endoplasmic reticulum
TGF-beta signaling pathway
N-Glycan biosynthesis
Notch signaling pathway
Lysine degradation
Endocytosis
Phosphatidylinositol signaling system
Dorso-ventral axis formation
Apoptosis - fly
Phototransduction - fly
p.adjust
0.01
0.02
0.03
0.04
Count
10
20
30

SUPPLEMENTARY MATERIALS (File S1)

Distinct gene expression dynamics in germ line and somatic tissue during ovariole morphogenesis in *Drosophila melanogaster*

Shreeharsha Tarikere, Guillem Ylla and Cassandra G. Extavour

These Supplementary Materials contain the following:

- Detailed Protocols
    I. Detailed protocol for staging larvae
    II. Detailed protocol for dissection and dissociation of larval ovaries
    III. Detailed RNA extraction protocol
    IV. Detailed library preparation protocol
- Key Resources Table
- Supplementary References
- Legends for Supplementary Tables S1 through S9 (this document)
- Supplementary Tables are provided in Supplementary Table files S1 through S9
- Supplementary Figures S1 through S13 with Legends (this document)

DETAILED PROTOCOLS

I. Detailed Protocol to stage larvae

1. Day 1: Collect 100 virgin females on the day before egg collection. Set the cross in a 50 ml media bottle with 50 males and leave at room temperature for 12h (overnight) to mate.
2. Make apple juice plates as follows:
    i. Boil 9g bacterial agar (Becton Dickinson catalog # 214050) in 300ml autoclaved distilled water.
    ii. Separately, dissolve and boil10g Sucrose in 100ml apple juice.
    iii. Mix the two solutions together while stirring with a magnetic stir bar.
    iv. Pour the media into 60x15mm plates once the temperature has cooled down to approximately 50°C.
    v. Cool plates without lids for two hours and then cover with lids and store inverted at 4°C.
    vi. These plates can be used for up to two weeks.
3. Day 2: Remove apple juice agar plates needed for each hour for up to eight hours and allow them to warm up to room temperature.
4. In a glass vial, place some yeast granules and add tap water to cover the granules. There must be a drop of water more than the yeast granules can soak up. This makes a paste of peanut butter consistency.
5. Using a steel spatula, smear a pea-sized amount of paste onto one end of a plate, for all the plates. Optimize this based on the number of flies such that the paste is neither completely consumed nor remains in excess after an hour-long collection.
6. Transfer the cross in the bottle to a 100 ml collection cage and cover it immediately with an apple juice egg-collection plate containing yeast smear. Fasten the setup with two rubber bands.
7. Incubate with the plate at the bottom for one hour at 25°C. All activity from this point until the point of dissection is done at 25°C. For the first hourly change, tap the bottom of the cage and quickly replace the old plate with a fresh collection plate.
8. Remove any flies stuck to the yeast patch with forceps, crush and discard them in the freezer.
9. After the final egg plate flip transfer the flies back into a bottle using a funnel. Discard the first two collection plates. Incubate all the remaining plates at 25°C.
10. Day 3: Start collecting larvae at the end of the second hour egg collection done the previous day. Collect uniformly sized larvae.
11. Transfer around 50 larvae from the same staged egg collection into a vial. Some yeast paste may be carried to the vial during this collection; try to keep this amount constant.
12. Collect until the final hour of the previous days' collection. Incubate the vials containing larvae at 25°C.

II. Detailed protocol for dissection and dissociation of larval ovaries

1. Begin dissections at the same time as that of the third plate from the egg collection day. Early-stage dissections take place three days from egg laying (72h AEL), middle (mid) stage four days from egg laying (96h AEL), and larval pupal stage (late) five days from egg laying (120h AEL).
2. Late stages are the easiest stage to locate and dissect, recognizable once the late third instar larvae have immobilized on the side of the vial and have a thickened cuticle. Use a fine wet paintbrush to dislodge them and place into cold 1x phosphate-buffered saline (PBS).
3. For early and middle stages, scrape the soggy layer of food from the vial using a spatula and spread it on a glass petri dish. Under a dissecting stereomicroscope select uniformly sized larvae, wash them in 1xPBS and place them into a fresh glass dish with 1xPBS.
4. Male larvae are easily identified by the two translucent spots that are the testes located at approximately 75% the length of the larval body from the anterior, and by their relatively smaller body size compared to females. Adjusting the light sources to be closer to the stage at the base of the glass dish allows better visualization of ovaries and testes. Discard the males by using forceps to transfer them onto a paper towel or kimwipe.
5. Using forceps, decapitate the larva. Gently squeeze out the inner contents from posterior to anterior using forceps.
6. Pull the larval body away gently with forceps from the fat-body/gut, while holding the body with another forceps. If done properly (after some experience) the larval body and the gut material separate from the fat body lobes. Well-fed and later stages of larvae are easier to dissect than younger stages.
7. The ovaries are located in the middle of the larval fat bodies. Ovaries are in a flower-like circular patch in the center of the fat body. They are around 100-500 μm in size depending on the stage, and appear as transparent tiny dots (early) to large dots (pupal) within the fat body.
8. Using two insulin needles, hold the fat body close to the ovary with one needle and use the other needle to cut closely around the ovary until it is released. Some areas of the fat body may remain initially, which will be removed later when the dissected ovaries are treated with trypsin.
9. Record the number of ovaries acquired in each batch.
10. Start to thaw an aliquot of liberase (stored at -20°C) to room temperature.
11. Dissect approximately 20 ovaries in this way, then place the glass dish on ice. Use a fresh dish for the next round of dissections.
12. Use a 12-well glass plate and add 200µl of trypsin to one well and the contents of the thawed liberase aliquot (~150µl) to the next well.
13. Clear any debris that surrounds the collected ovaries into a separate glass dish using a needle. Using a 2µl pipette filter tip pre-rinsed in trypsin solution, transfer the dissected ovaries into the trypsin well.
14. After transferring the last batch of ovaries into trypsin solution, incubate for 10 minutes in trypsin solution. Meanwhile label two RNA lobind tubes (Eppendorf 8077-230), with a black waterproof marker indicating the genotype of the

dissection, the date and the number of ovaries. Add the desired amount (100-200µl) of Trizol (based on the anticipated number of cells after sorting) to one of the two tubes (tube #1) – this tube will be used to collect cells following FACS (step 18). Leave the second tube (tube #2) empty – this will be used to collect the dissociated cells following enzymatic treatment (step 17). Place both tubes on ice.

15. Using a 2µl filter tip pre-rinsed in liberase solution, transfer the ovaries from trypsin into liberase. Dissociate them using needles until no clusters of cells can be seen. Incubate them in liberase solution for ten minutes, starting the clock when the ovaries are transferred from trypsin (includes time needed to dissociate using needles).
16. Place 200ml of liquid nitrogen in a liquid nitrogen container.
17. Using a 200µl pipette filter tip pre-rinsed in 1xPBS, pipette the tissue in liberase up and down gently ten times to dissociate and resuspend the cells. Transfer the contents to tube #2 and place it on a microtube vortexer for one minute. Meanwhile rinse the well that contained liberase with 1.4 ml of 1xPBS and pipette up and down 10 times. Add this volume into the tube and vortex for an additional 10 minutes.
18. Place the tube on ice, remove one glove (so you can safely touch door handles) and carry the tube on ice, and your liquid nitrogen container, to your FACS facility/machine.
19. Collect cells following FACS into the Trizol in tube #1.

III. Detailed RNA extraction protocol

1. Because RNA extraction from FACS-sorted samples involves precious samples, take care to work in an RNase-free environment. Wear lab coat, gloves and safety goggles while working with Trizol and handling RNA samples.
2. Clean the table, ice bucket, pipettes, centrifuge, pellet pestle motor, table top vortexer and tabletop mini-vortex first with 70% ethanol and then with RNase zap (Thermo Fisher AM9780) on tissue paper.
3. Pellet pestles are cleaned first in 100% ethanol and then in nuclease free water (Thermo Fisher 10977015). Prior to use the pestles should be sterilized in a glass beaker covered in aluminum foil by autoclaving in a liquid cycle for 30 min.
4. Add 500 µl Trizol into a lo-bind tube, which will be used to pre-rinse the pellet pestle before crushing each sample.
5. Thaw Trizol cell samples (retrieved in tube #1 at step 19 in Protocol II) at room temperature (RT) and place them on ice. Spin in tabletop mini-vortex for 10 seconds.
6. Crush each cell sample with a separate pellet pestle pre-rinsed in the Trizol set aside for this purpose in step 4.
7. Crush cells in Trizol with a pre-rinsed pellet pestle. Add an equal volume of 100% ethanol, mix with pestle and vortex briefly. Spin samples briefly in tabletop mini-vortex.
8. Pipette the sample onto a zymo spin column.
9. Wash column in 400 µl RNA wash buffer. Centrifuge at 10,000 g for 1 minute at RT.
10. Thaw DNase (Zymo kit: 6unit/µl) from storage at -20°C. Mix 5 µl DNase with 35 µl DNase digestion buffer per sample. Add 40 µl of DNase mix to each column.
11. Incubate at RT for 15 minutes. Centrifuge at 10,000 g for 1 minute at RT..
12. Add 400 µl RNA pre-wash buffer to the column and centrifuge. Repeat this step one more time.
13. Add 700 µl of RNA wash buffer. Centrifuge twice to completely remove the buffer from the column, each time at 10,000 g for 1 minute at RT..
14. The column-bound RNA is eluted in two steps using nuclease-free water. Add 25 µl water and incubate for 15 minutes at RT. Centrifuge at 10,000 g for 1 minute at RT. Add another 20 µl of water to elute a second time.
15. The library preparation protocol below (**IV**) requires RNA in a 50 µl volume; an additional 5 µl is reserved for quality control measurements.
16. Quantify RNA first in Nanodrop RNA-40 measurement with 1.5 µl of the sample. Then use Qubit high sensitivity RNA kit to quantify 1µl of the sample (10µl Standard1/2+190 Buffer/Reagent, 1 µl sample+199 Buffer/Reagent).
17. Use a high-sensitivity RNA tape to quantify and measure the RNA integrity (RIN) in a Tapestation. Use the standard protocol for high-sensitivity RNA quantification using an electronic ladder for estimation of size.

IV. Detailed library preparation protocol (Takara Apollo system)

This is the Wafergen/Takara protocol. Low-throughput protocols process a maximum of eight samples at a time in the Apollo liquid handling unit. Label the individual tubes in the strip with the respective sample identifiers and make sure the marked wells are in the same 1-8 sequence direction at all times. Use the protocol images to ensure accurate placement of strip tubes and double check tube placements regularly. RNA extraction, Poly A selection and Library preparation should be done on the same day.

IV.i. Poly A selection protocol

1. Verify that the volumes of RNA samples are at least 50 µl using micropipettes and pipette them into a strip tube (Fisher scientific 14-222-252). Label tube as "RNA samples". Spin down in a tabletop mini-centrifuge and keep on ice until step 7.
2. Wipe the inner surface of the Apollo system with RNase zap (Thermo Fisher AM9780) and 70% Ethanol.
3. Ensure the trash container is empty. The machine shows an error when trash is full.
4. Cool the Apollo machine to 4°C using the standard protocol 'cooling' function.
5. Place empty reservoirs (Apollo 640087) in Block-6 row 1 and 2.
6. Fill Block-5 row 1,2 and 3 with filter tips (Apollo 640084).
7. Place a new microplate in Block-2.
8. In two new reservoirs, add 10 ml (protocol says 4/5ml) of Reagent 1 and place it in row 1 and 10ml of reagent 2 in row 2.
9. Place empty 8 strip tubes in Block-3 row 4 and 5 and in Block-4 row 1, 2 and 3.
10. Label an empty strip tube as "products" and place it in Block-3 row 8. This strip tube will contain the poly-A selected mRNA at the end of this run. Label each tube with the sample names.
11. Aliquot 80µl of Reagent 3 (it is a clear solution) into an 8 well strip tube into wells corresponding to the RNA sample. Label as "Reagent 3" to distinguish it from RNA samples.
12. Place Reagent 3 strip tube into Block-3 row 3.
13. Gently pipette the magnetic bead reagent 4 until it is uniformly resuspended. Aliquot 15 µl into the same number and locations of the strip tube containing RNA sample.
14. Place the RNA sample tube in Block-3 row 1.
15. At last place the magnetic bead reagent 4 in Block-3 row 2. Place retainers for Blocks 3 and 4 and lock them.
16. Restart the machine by switching off and on. On the touchscreen navigate to the latest version of PolyA8 protocol ("User Maintenance" > "PrepX_PolyA8_betaV2").
17. The run lasts for 45 minutes, during which time the reagent mixes for the subsequent library preparation protocol (**IV.ii**) should be set up.
18. Remove strip tubes while checking for uniform volumes. Note any discrepancies.

19. Check the volume of the “product” tube, cap it and spin it in a minispin. The volume should be approximately 19µl. Cap and store on ice until the next step. This mRNA “product” strip tube will be used as “sample” in the cDNA library preparation step.

IV.ii. PrepX mRNA8 Library preparation protocol.

1. Prepare the RNase III mix and Reverse Transcription (RT) reaction mix for the number of mRNA samples and an additional sample.
2. In a RNase free tube mix 2µl each of RNase Buffer III and RNase III enzyme (Thermo Fisher 18080093) per sample required. Pipette gently and give a brief spin on a tabletop mini-centrifuge. Leave the tube on ice.
3. Mix the following reagents per reaction to make the RT reaction mix and place it on ice:
    - 5X First strand buffer 16µl
    - 0.1M DTT 08µl
    - dNTP 04µl
    - Superscript III Reverse Transcriptase 02µl
    - Murine RNase inhibitors 01µl

Setting the apollo system blocks:

4. Place empty strip tubes in Block-3 rows 1 and 2 and in Block-4 row 6.
5. Label a strip tube as products and place it in Block-3 row 5.
6. Fill filter tips in Block-5 rows 1-7, fill black piercing tips (Apollo 640085) in row 12.
7. Fill 1.1 ml tube strips in Block-1 row 1,2 and 3.
8. Place a used microplate in Block-2.
9. Cool the Apollo machine to 4°C using standard protocol ‘cooling’ function.
10. Place empty reservoirs in position 2 and 4 of Block-6. Add 15 ml of 100% molecular grade ethanol (Sigma E7023) in reservoir 3 and 15 ml of nuclease free water (part of Takara library prep kit 640096) in reservoir 1.
11. In a new strip tube, aliquot 4 µl of RNase III mix into each tube corresponding to the sample and place it in Block-4 row 5.
12. In a new strip tube, aliquot 31 µl of RT reaction mix into each corresponding sample tube and place it in Block-4 row 7.
13. Gently pipette A-line beads and aliquot 200µl into a fresh strip tube at the corresponding wells to that of the sample and place it in Block-4 row 8.
14. Thaw blue enzyme and orange adapter/primer strips of the PrepX mRNA8 kit (Takara 640096) on ice. The number of these strips is the same as the sample number. Flick the bottom of tubes to dislodge liquid and mix uniformly. Spin down and place it back on ice.
15. Sometimes a solid precipitate might be visible in the enzyme strip tubes. It generally dissolves after flicking. Do not use the strip if it is not soluble after thawing and mixing.
16. Place the Blue strip tubes with the arrow pointing up in Block4 rows 9-12 in columns corresponding to the mRNA samples.

17. Place the orange strip tubes in Block-4 from row 1-4 in columns corresponding to the mRNA sample tubes.
18. Verify the filter and piercing tips in Block-5, 1.1ml tubes in Block-1, mock microplate in Block-2, Reservoir 3 with 100% ethanol and Reservoir-1 with Nuclease free water, Reservoirs 4-5 empty place holders, three sets of empty tubes (one for cDNA products) and mRNA sample tube in Block-3 and in Block-4 Blue and orange strips, one empty strip tube, RT, RNase III and bead strip tubes in specified locations.
19. Lock Block-3 and 4 with retainer plates.
20. In the touchscreen control select User maintenance > Run the protocol 'PrepX_mRNA8_200bp_BetaV1.scb. The screen does not show any progress bar. This program runs for 5 hours.
21. cDNA after this step is fairly stable and can be processed the next day. Keep at 4°C until processing.

IV.iii. PCR Amplification of the Libraries.

1. Check the volume of the "product" tube, cap it and spin it in a minispin. The volume should be around 19µl. Cap and store on ice until the next step.
2. Prepare PCR master mix with 25µl long Amp Taq (NEB M0323S) and 2.5 µl of SR primer per sample on ice.
3. Add unique index primers (PrepX RNAseq index 1-48) to each of the cDNA products and add the PCR master mix. Make up the total volume in each tube to 50µl with Nuclease free water.
4. Mix well and spin down. Place the strip tube in a PCR machine and run a 15 cycle PCR amplification reaction as shown below:

Temp	94℃	94℃	60℃	65℃	65℃	10℃
Time	60 sec	30 sec	30 sec	30 sec	7 min	hold
Cycles		---------- 14 x ----------			extension	

6. After completion of PCR cool the samples and spin down.
7. PCR Cleanup using the Apollo system PrepX_PCR clean up8 protocol.
8. Cool the Apollo machine to 4°C using standard protocol 'cooling' function.
9. Place filter tips in Block-5 row 1. Block-6 is similar to the library preparation protocol, empty reservoirs in position 2 and 4 of Block-6. 10 ml of 100% molecular grade ethanol (Sigma E7023) in reservoir 3 and 15 ml of nuclease free water (part of Takara library prep kit 640096)) in reservoir 1.
10. Place empty strip tubes in Block-3 row 2 and 3. Label a strip tube as "cleaned cDNA product" and place it in row 4. Place the amplified cDNA samples in row 1.
11. Aliquot 50 µl of A-line beads into a fresh strip tube at corresponding locations to the samples. Place this tube in the end of the set up.
12. Place retainers on Block-3 and 4 and lock it.

13. Run Utility apps> PCR cleanup 8 protocol on touch screen. The run lasts 20 minutes.
14. Check volumes of the cleaned ‘product’ tube. Spin down and place on ice.
15. Run Qubit to quantify the cDNA with the high sensitivity DNA kit. Use a Tapestation to run the gel and measure cDNA quantity using ladder and high sensitivity tape (Agilent 5067-5579). Depending on the qubit quantification, higher concentrations use DNA 1000 tape (Agilent 5067-5582).
16. Transfer the cDNA in strip tubes to a lobind RNA tube (Eppendorf 8077-230) and label them with details about the sample-genotype, date of cDNA prep, it is a cDNA library, the concentration of the sample, and the volume of sample remaining after quantification. Store at -80°C until it is ready to be sequenced.
17. Calculate the total molar concentration of the lane and the dilution needed for each library to make it equimolar. Mix the volumes in a single lobind tube and submit to the sequencing facility.

326 **KEY RESOURCES TABLE**
327

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals		
Hoechst 33342	Thermo Fisher	Cat# H1399
Dulbecco's Phosphate Buffered Saline PBS	Thermo Fisher	Cat# 14190144
Trypsin 0.25%	Thermo Fisher	Cat# 25200056
Liberase 2.5%	Sigma	Cat# 5401119001
Trizol	Thermo Fisher	Cat# 15596206
Ethanol molecular 200 grade	Sigma	Cat# E7023
Nuclease free water	Thermo Fisher	Cat# 10977015
Magnetic Beads	A-line	Cat# C1003
Triton X100	VWR	Cat# 97062-208
Critical Commercial Assays		

Zymo RNA Micro-prep kit	Zymo Research	Cat# R2060
Superscript III Reverse Transcriptase	Thermo Fisher	Cat# 18080093
Takara PrepX PolyA-8	Takara	Cat# 640098
Takara PrepX mRNA-8	Takara	Cat# 640096
Qubit RNA HS Assay Kit	Thermo Fisher	Cat# Q32852
Qubit DNA HS Assay Kit	Thermo Fisher	Cat# Q32854
LongAmp Taq DNA Polymerase	New England BioLabs	Cat# M0287S
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE172015
Experimental Models: Organisms/Strains		
D. melanogaster: bab1 GAL4: *w[*]; P{w[+mW.hs]=GawB}bab1[PGAL4-2]/TM6B, Tb[1]*	Bloomington Drosophila Stock Center	BDSC:6803; FlyBase:ID FBst0006803
D. melanogaster: nos GAL4: *P{w[+mC]=UAS-Dcr-2.D}1, w[1118]; P{w[+mC]=GAL4-nos.NGT}40*	Bloomington Drosophila Stock Center	BDSC:25751; FlyBase: ID FBst0025751

D. melanogaster: *w1118, P{UAS Stinger}*	(BAROLO *et al.* 2000)	UAS Green Stinger on X
Instruments		
MoFlo Astrios EQ Cell sorter	Beckman Coulter	B25982
Motorized pellet pestle	Kimble	Cat# 749540-0000
NanoDrop	Nanodrop	ND1000
Qubit 3.0 Fluorometer	Thermo Fisher	Cat# Q33216
Tapestation	Agilent	2200/4200
PCR Thermal cycler	Bio-Rad	C1000
Illumina Hi Seq	Illumina	2500
Consumables		
Insulin Syringe	Becton Dickinson	Cat# 328418
RNA lo-bind tubes	Eppendorf	Cat# 8077-230
High Sensitivity RNA ScreenTape	Agilent	Cat# 5067-5579

High Sensitivity DNA ScreenTape	Agilent	Cat# 5067-5584
DNA 1000 ScreenTape	Agilent	Cat# 5067-5582
Axygen PCR 8-strip tubes	Fisher Scientific	Cat# 14-222-252
Apollo Filter tips 300027	Takara	Cat# 640084
Apollo Piercing tips 300028	Takara	Cat# 640085
Apollo Reservoirs 300031	Takara	Cat# 640087
Software and Algorithms		
Drosophila melanogaster genome version	(LARKIN *et al.* 2021)	Dmel_r6.36_FB2020_05
RSEM v 1.3.3	(LI AND DEWEY 2011)	https://deweylab.github.io/RSEM/
STAR aligner v 2.7.6a	(DOBIN *et al.* 2013)	https://github.com/alexdobin/STAR
DESeq2 v 1.26.0	(LOVE *et al.* 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Bcl2fastq2 v2.20		https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html

SUPPLEMENTARY REFERENCES

Barolo, S., L. A. Carver and J. W. Posakony, 2000 GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. Biotechniques 29**:** 726, 728, 730, 732.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29**:** 15-21.

Larkin, A., S. J. Marygold, G. Antonazzo, H. Attrill, G. Dos Santos *et al.*, 2021 FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res 49**:** D899-D907.

Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12**:** 323.

Love, M. I., W. Huber and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15**:** 550.

Slaidina, M., T. U. Banisch, S. Gupta and R. Lehmann, 2020 A single-cell atlas of the developing Drosophila ovary identifies follicle stem cell progenitors. Genes Dev 34**:** 239-249.

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table S1: RNA-seq sample metadata, including the sample name, biological and technical replicate information, tissue, stage, cDNA concentration, PrepX Index used, number of ovaries used, number of cells counted by FACS (not applicable to whole ovary samples), number of raw reads, aligned reads, and percentage of aligned reads as reported by the RSEM summary.

Supplementary Table S2: Differentially expressed genes (padj<0.01) between the consecutive developmental stages of whole ovary libraries. The column “Contrast” indicates whether the gene was found differentially expressed in early vs mid stages or mid vs late stages, and the column “Up_in” indicates which library the gene was overexpressed in.

Supplementary Table S3: Differentially expressed genes (padj<0.01) at each stage compared to the other two stages of the whole ovary. The column “Stage_up” indicates which stage the given gene was overexpressed in.

Supplementary Table S4: Differentially expressed genes (padj<0.01) between germ cells and somatic cells at all studied stages. The column “Up_in” indicates whether the gene was found upregulated in germ cells or somatic cells.

Supplementary Table S5: Differentially expressed genes (padj<0.01) between germ cells and somatic cells at each individual stage. The column “Up_in” indicates whether the gene was found upregulated in germ cells or somatic cells, and the column “Stage” indicates the stage (early, mid or late) in which the test was performed.

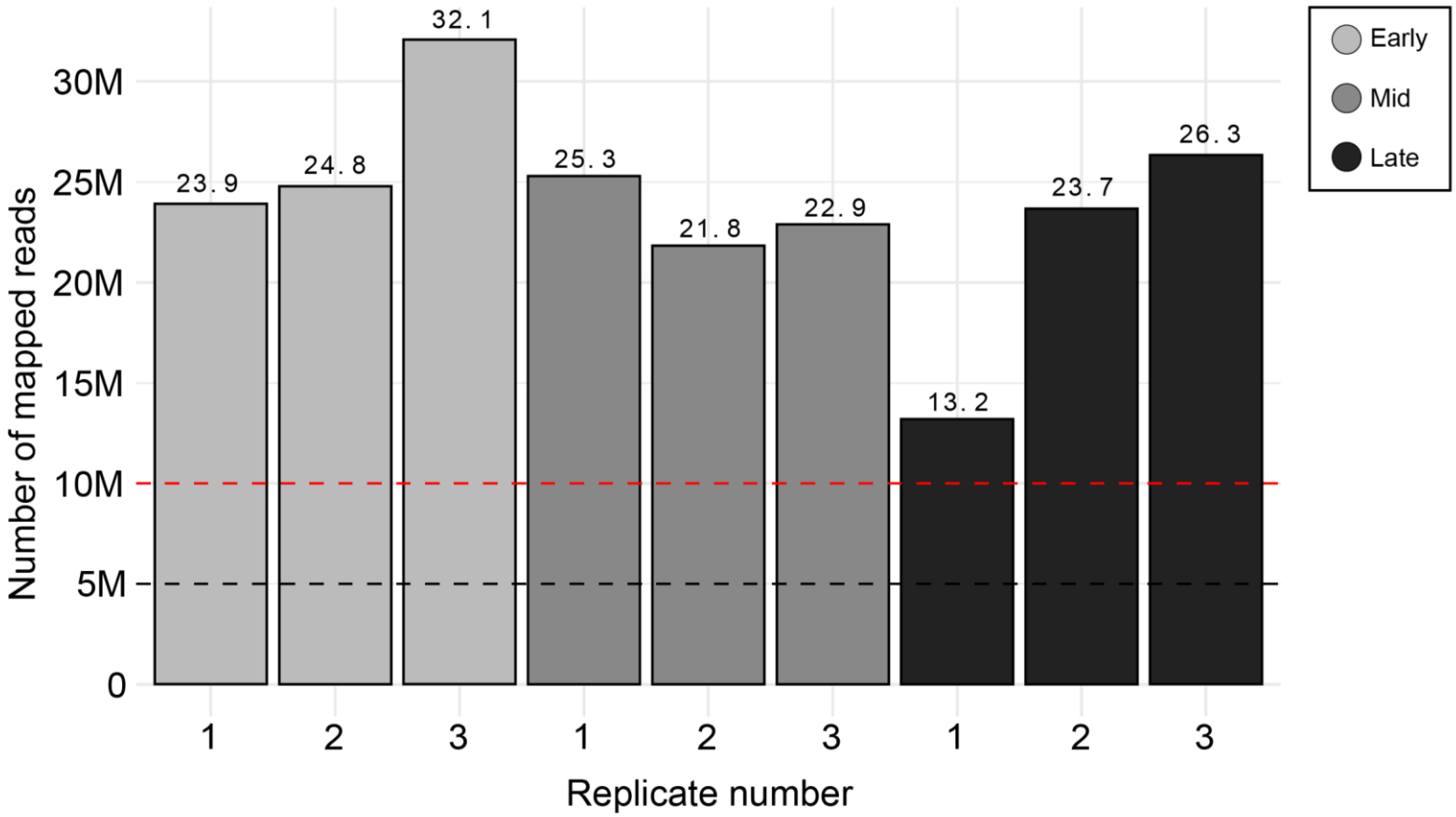
Supplementary Table S6: Differentially expressed genes (padj<0.01) between the consecutive developmental stages of somatic cell libraries. The column “Transition” indicates whether the gene was found differentially expressed in the transition from early to mid-stage or from mid to late stage, and the column “Up_Down” indicates whether the gene was up or down regulated in the given transition.

Supplementary Table S7: Differentially expressed genes (padj<0.01) at each stage compared to the other two stages of the somatic tissue library. The column “Stage_up” indicates the stage that a given gene was found overexpressed at.

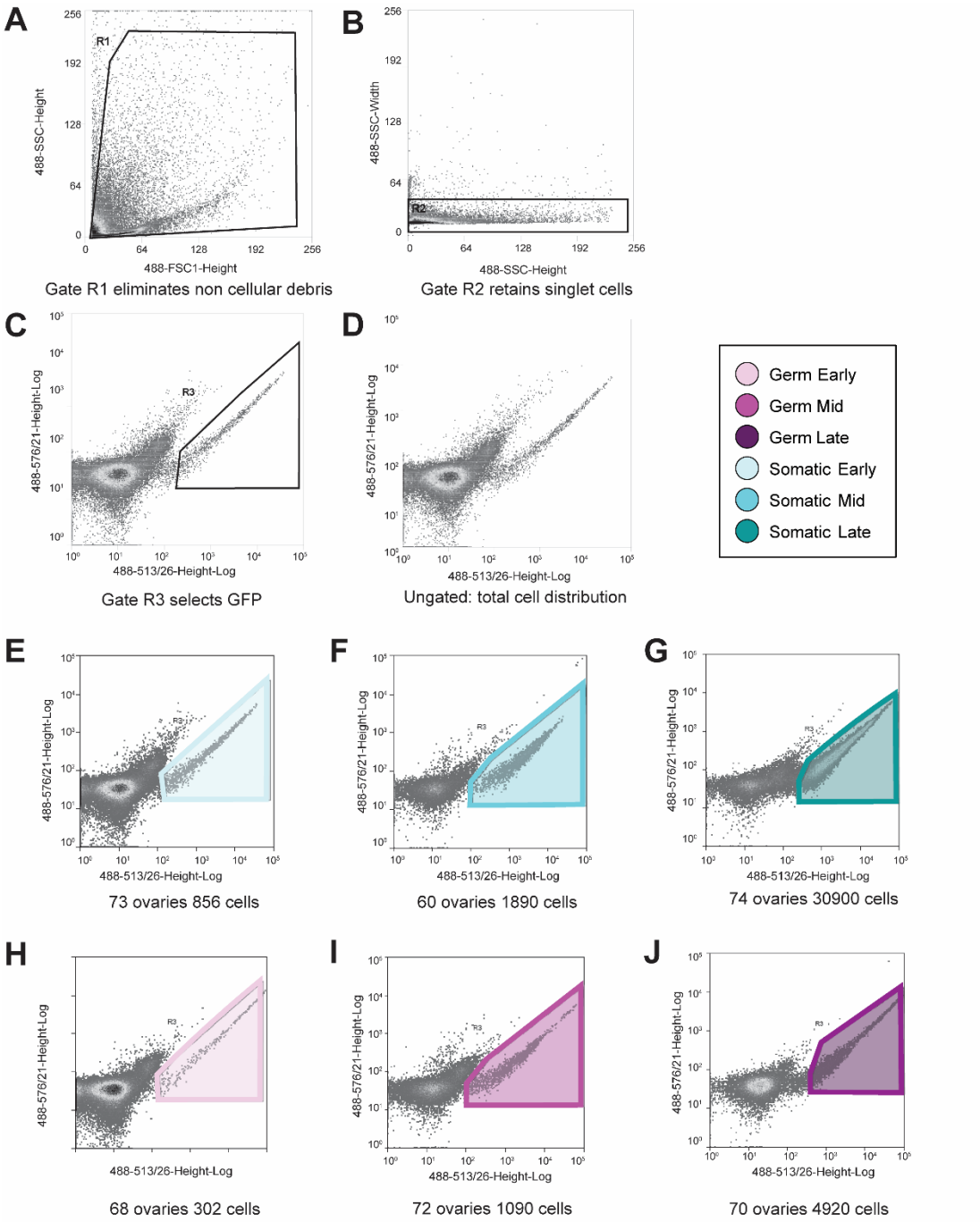
Supplementary Table S8: Differentially expressed genes (padj<0.01) between the consecutive stages of germ cell libraries. The column “Transition” indicates whether the gene was found differentially expressed in the transition from early to mid-stage or from mid to late stage, and the column “Up_Down” indicates whether the gene is up or down regulated in the given transition.

Supplementary Table S9: Differentially expressed genes (padj<0.01) at each stage compared to the other two stages of the germ cell library. The column “Stage_up” indicates the stage at which the given gene was found to be overexpressed.

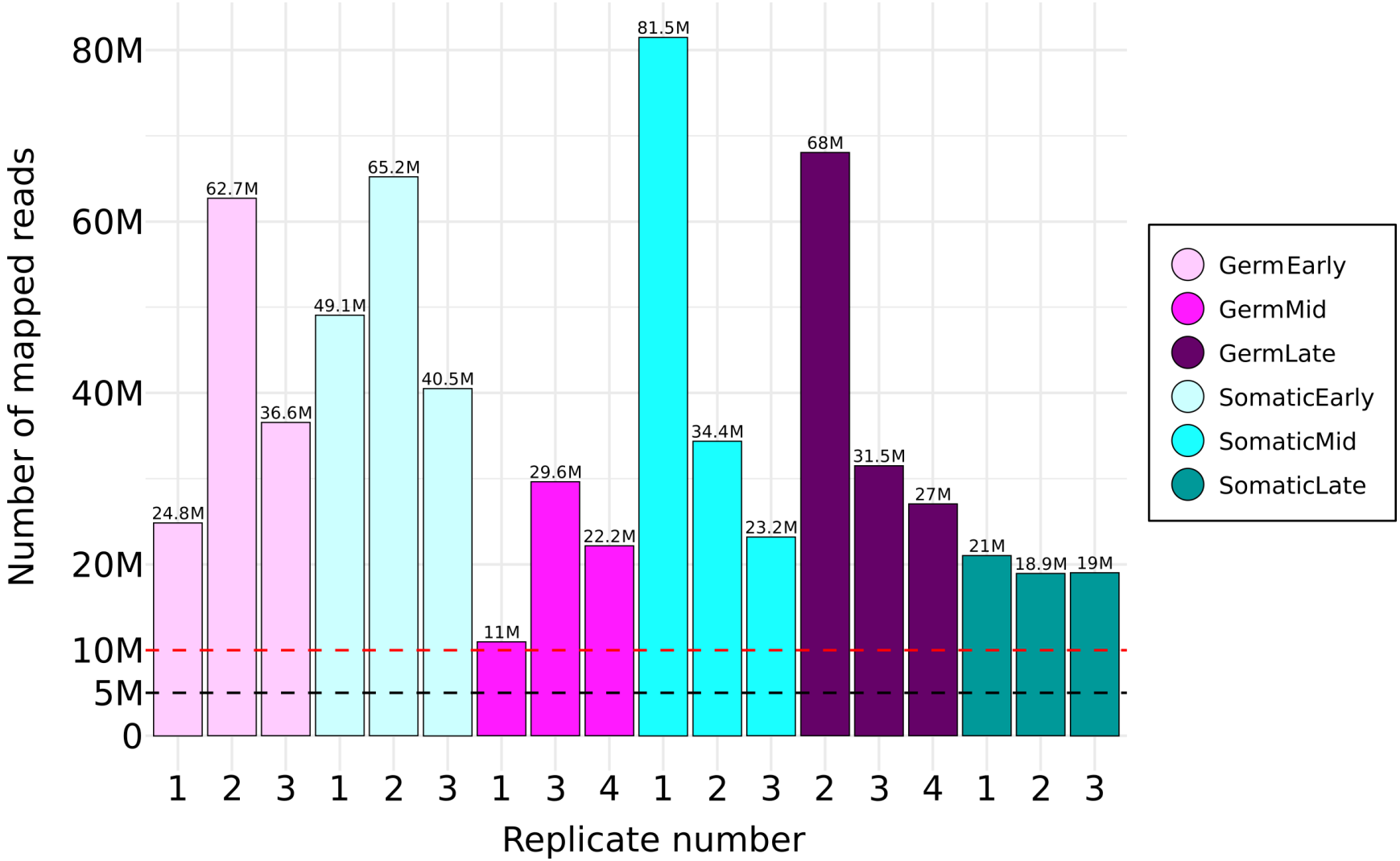
SUPPLEMENTARY FIGURES AND LEGENDS



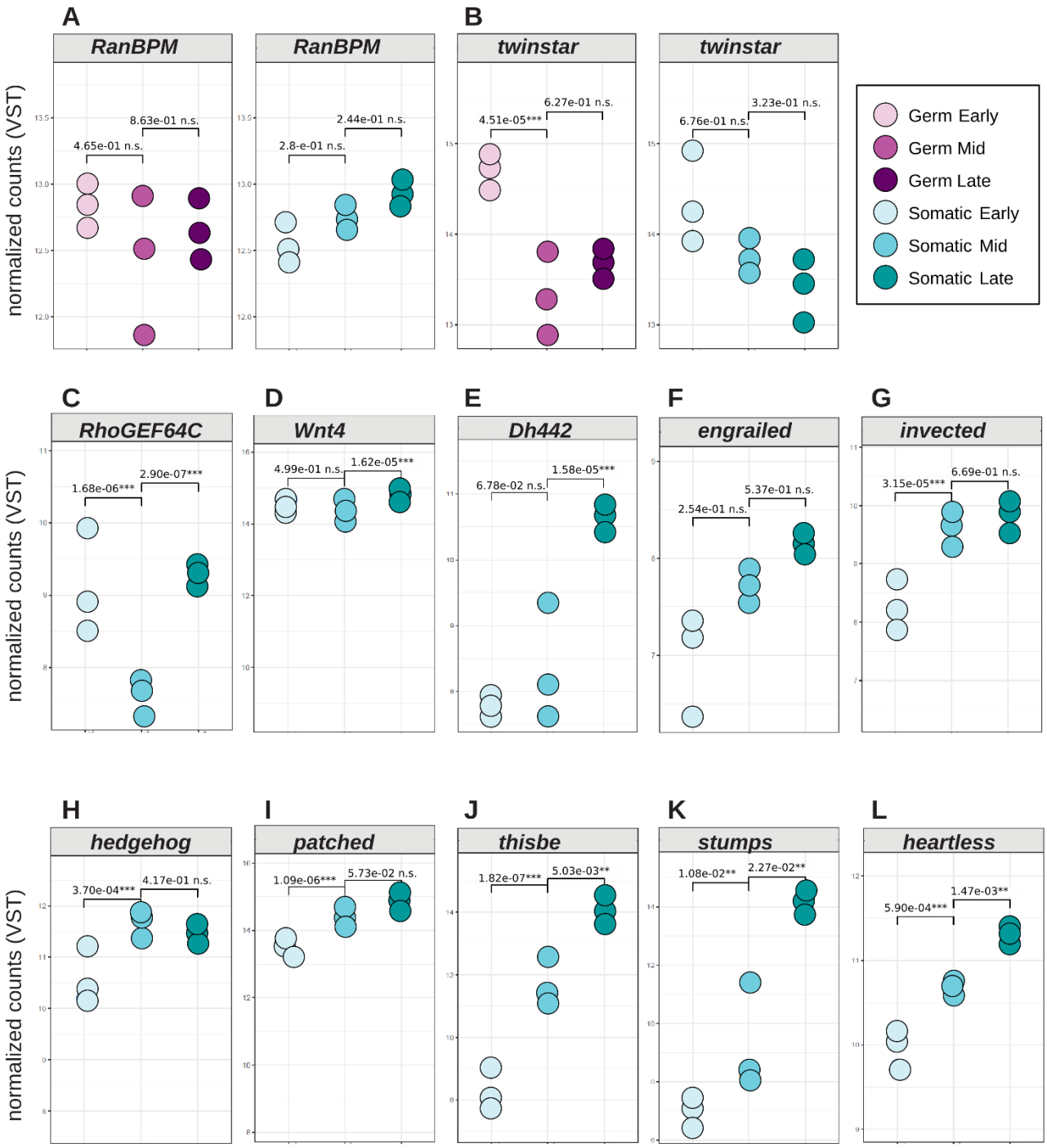
Supplementary Figure S1. Number of aligned reads in each of three biological replicates of whole ovary RNA-seq samples. The Black dashed line: five million reads; red dashed line: ten million reads.



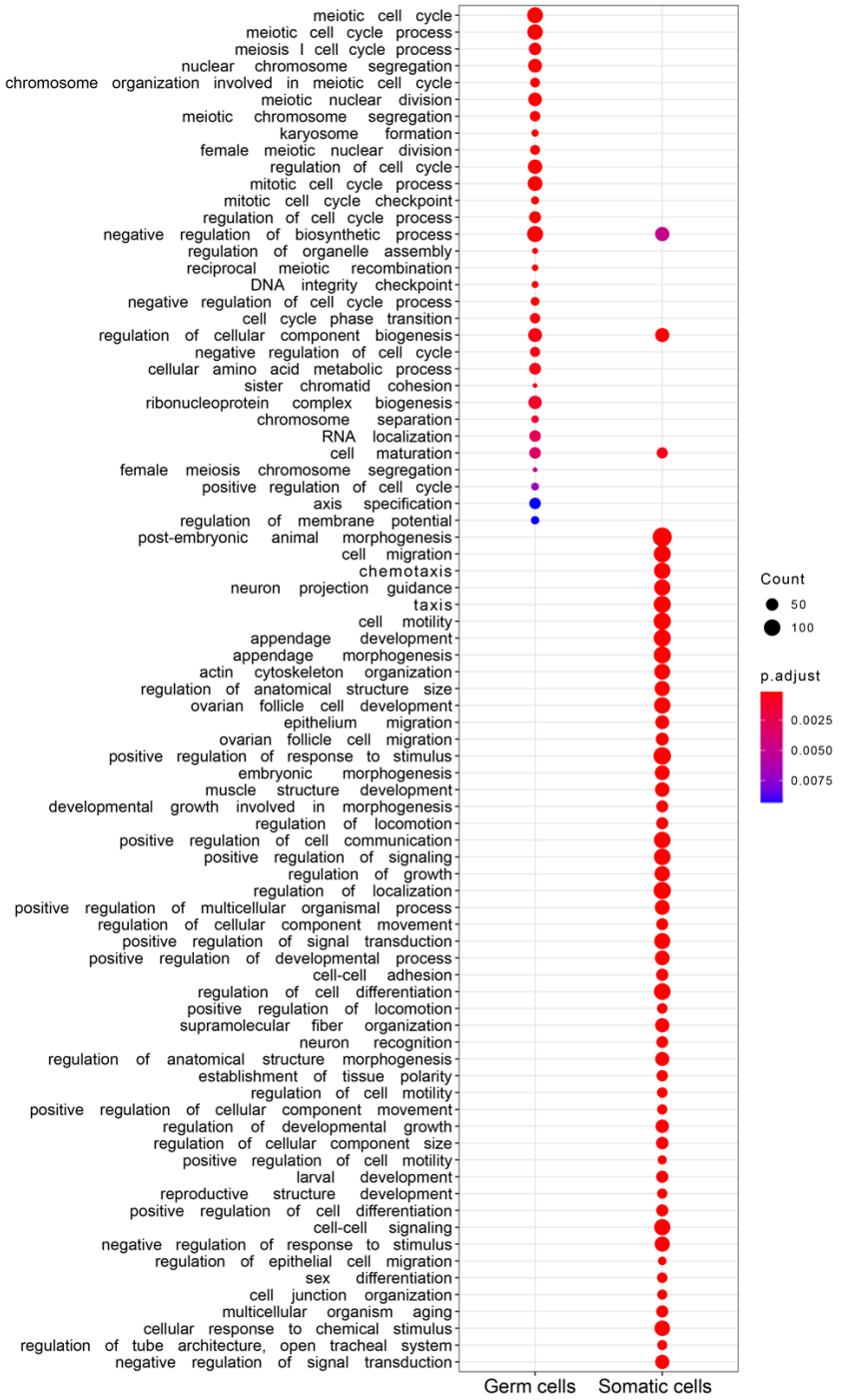
Supplementary Figure S2. Representative FACS cell density plots of GFP-positive cell sort from dissociated ovaries showing gating scheme used to obtain GFP-positive singlet cell population. A) Elimination of cellular debris using R1 (cells out of R1) gate. B) Elimination of non-singlets (cells outside R2) using R2 gate. C) Selection of GFP-positive cells through R3 (cells inside R3) gate. D) Ungated plot showing distribution of all cells. E-J) Representative R3 gated plots showing number of GFP-positive cells for similar number of ovaries.



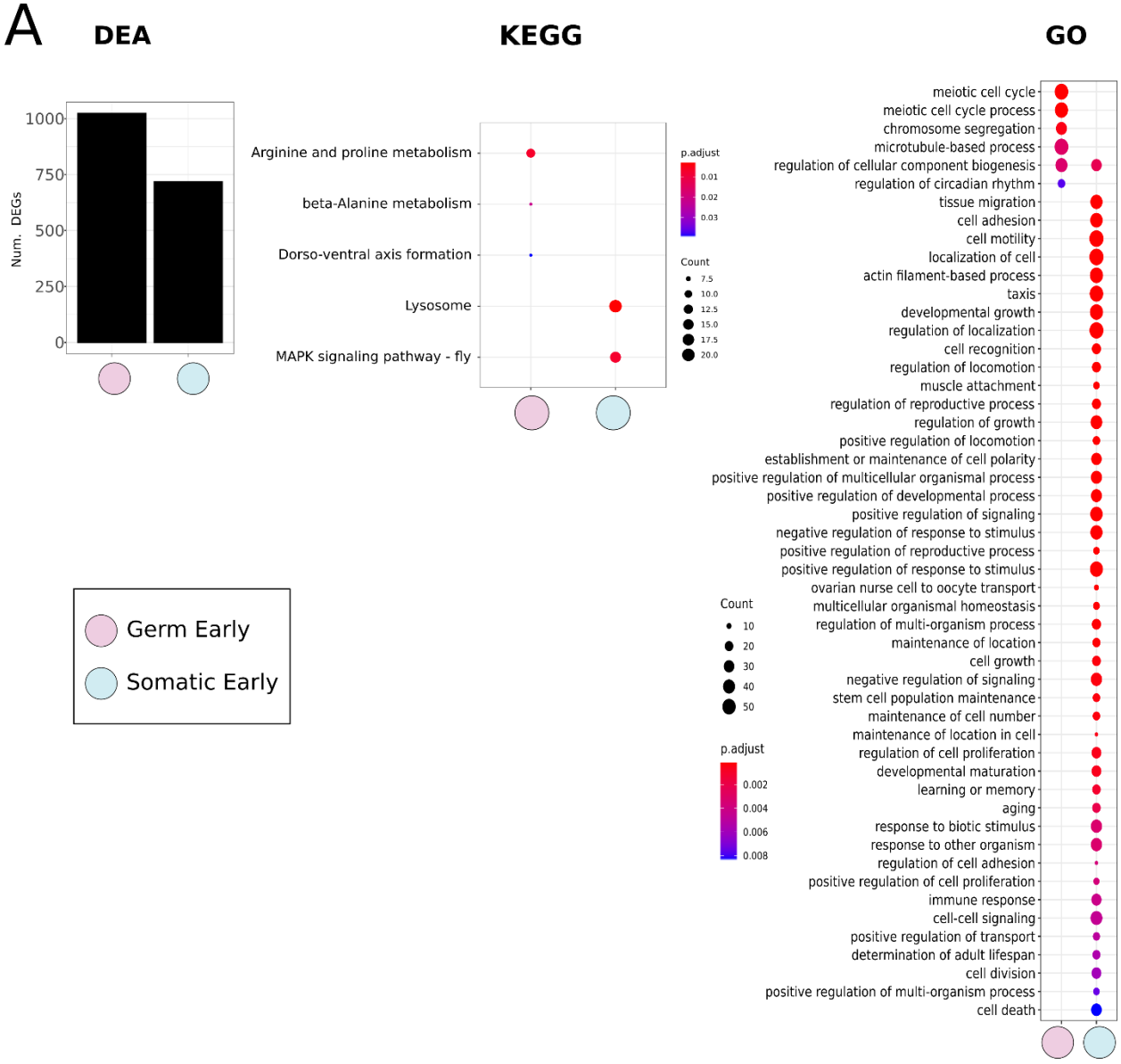
Supplementary Figure S3. Number of aligned reads in each of three tissue-specific biological replicate samples used for the analyses presented in this study. Black dashed line: five million reads; red dashed line: ten million reads.

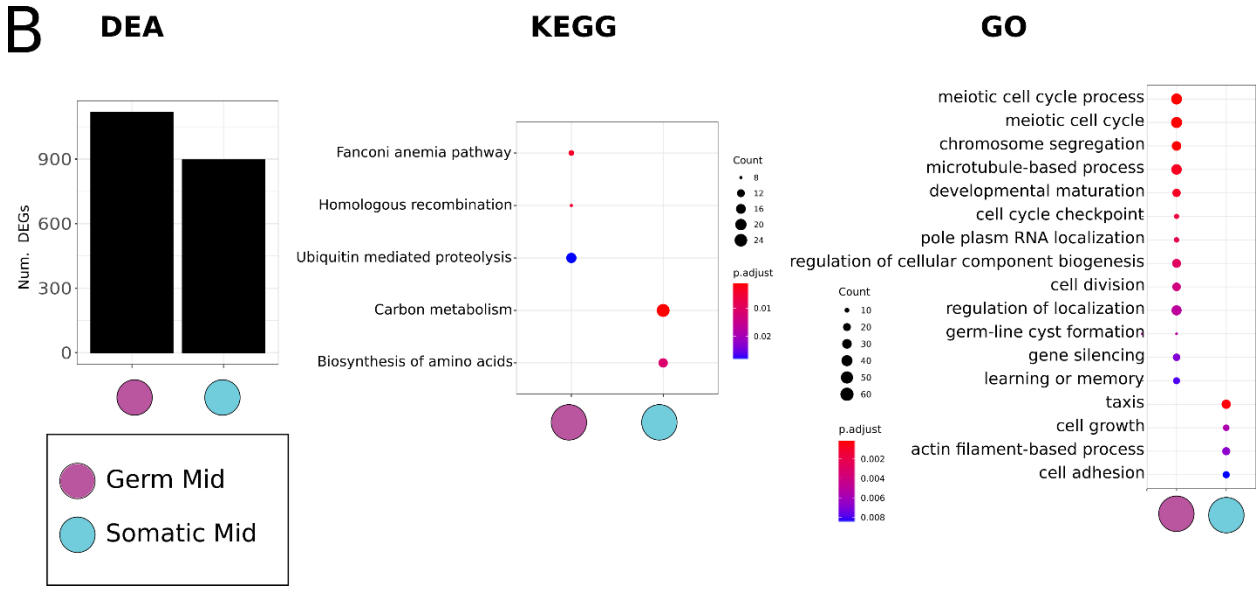


Supplementary Figure S4. Dot plots of genes differentially expressed across stages. Expression in VST counts of genes in each tissue-specific RNA-seq library. The adjusted p-values shown were calculated in the differential expression analysis with DESeq2. *p-value<0.05, **p-value<0.01, ***p-value<0.001, n.s. not significant

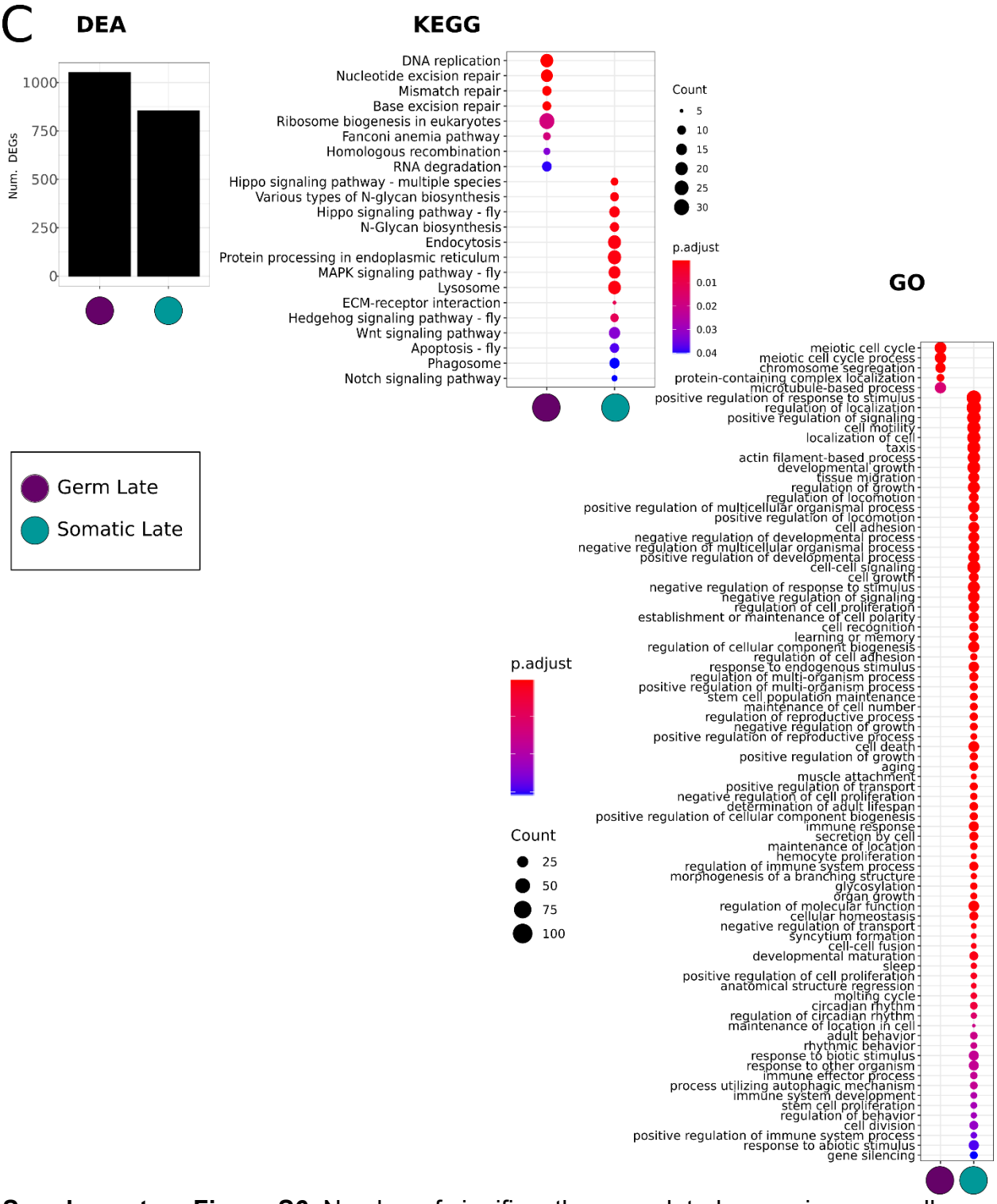


Supplementary Figure S5. Gene Ontology terms (GO-terms) for Biological Process level 4 that were significantly enriched (BH adjusted p-value<0.01, minimum number genes with the term=30) within the set of genes differentially expressed (BH adjusted p-value<0.01) between germ cells and somatic cells. The circle size is proportional to the number of genes with the GO-term in the corresponding gene set. The color of the circle indicates the adjusted p-value.

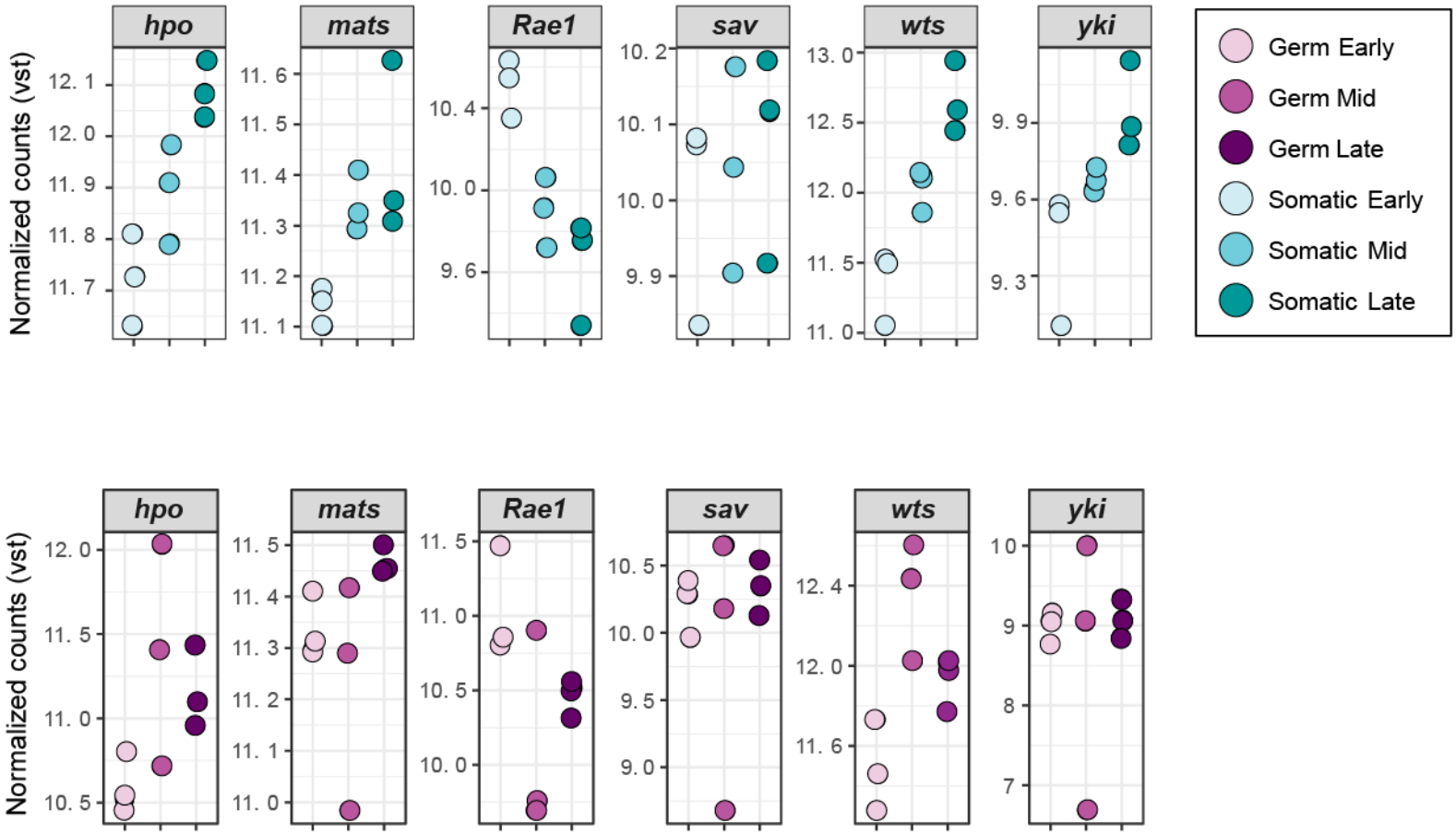

A
DEA
KEGG
GO
Num. DEGs
1000
750
500
250
0
Arginine and proline metabolism
beta-Alanine metabolism
Dorso-ventral axis formation
Lysosome
MAPK signaling pathway - fly
p.adjust
0.01
0.02
0.03
Count
7.5
10.0
12.5
15.0
17.5
20.0
meiotic cell cycle
meiotic cell cycle process
chromosome segregation
microtubule-based process
regulation of cellular component biogenesis
regulation of circadian rhythm
tissue migration
cell adhesion
cell motility
localization of cell
actin filament-based process
taxis
developmental growth
regulation of localization
cell recognition
regulation of locomotion
muscle attachment
regulation of reproductive process
regulation of growth
positive regulation of locomotion
establishment or maintenance of cell polarity
positive regulation of multicellular organismal process
positive regulation of developmental process
positive regulation of signaling
negative regulation of response to stimulus
positive regulation of reproductive process
positive regulation of response to stimulus
ovarian nurse cell to oocyte transport
multicellular organismal homeostasis
regulation of multi-organism process
maintenance of location
cell growth
negative regulation of signaling
stem cell population maintenance
maintenance of cell number
maintenance of location in cell
regulation of cell proliferation
developmental maturation
learning or memory
aging
response to biotic stimulus
response to other organism
regulation of cell adhesion
positive regulation of cell proliferation
immune response
cell-cell signaling
positive regulation of transport
determination of adult lifespan
cell division
positive regulation of multi-organism process
cell death
Count
10
20
30
40
50
p.adjust
0.002
0.004
0.006
0.008
Germ Early
Somatic Early


B
DEA
KEGG
GO
Num. DEGs
900
600
300
0
Germ Mid
Somatic Mid
Fanconi anemia pathway
Homologous recombination
Ubiquitin mediated proteolysis
Carbon metabolism
Biosynthesis of amino acids
Count
p.adjust
0.01
0.02
meiotic cell cycle process
meiotic cell cycle
chromosome segregation
microtubule-based process
developmental maturation
cell cycle checkpoint
pole plasm RNA localization
regulation of cellular component biogenesis
cell division
regulation of localization
germ-line cyst formation
gene silencing
learning or memory
taxis
cell growth
actin filament-based process
cell adhesion
0.002
0.004
0.006
0.008

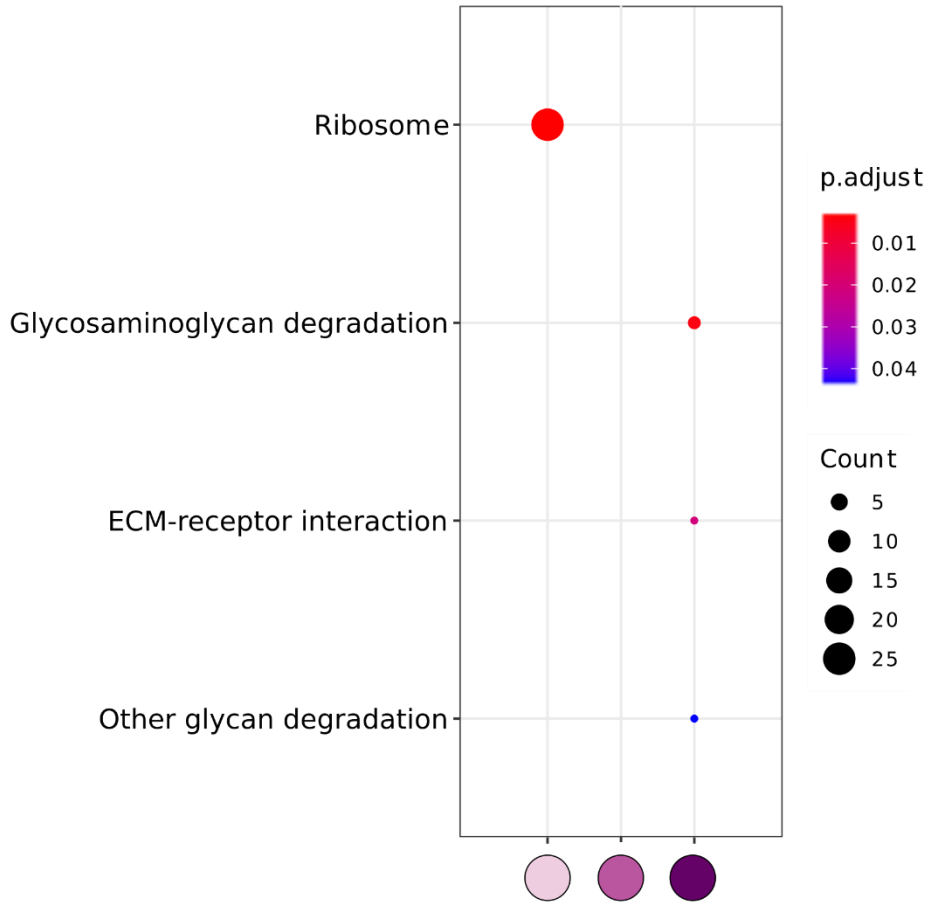
432



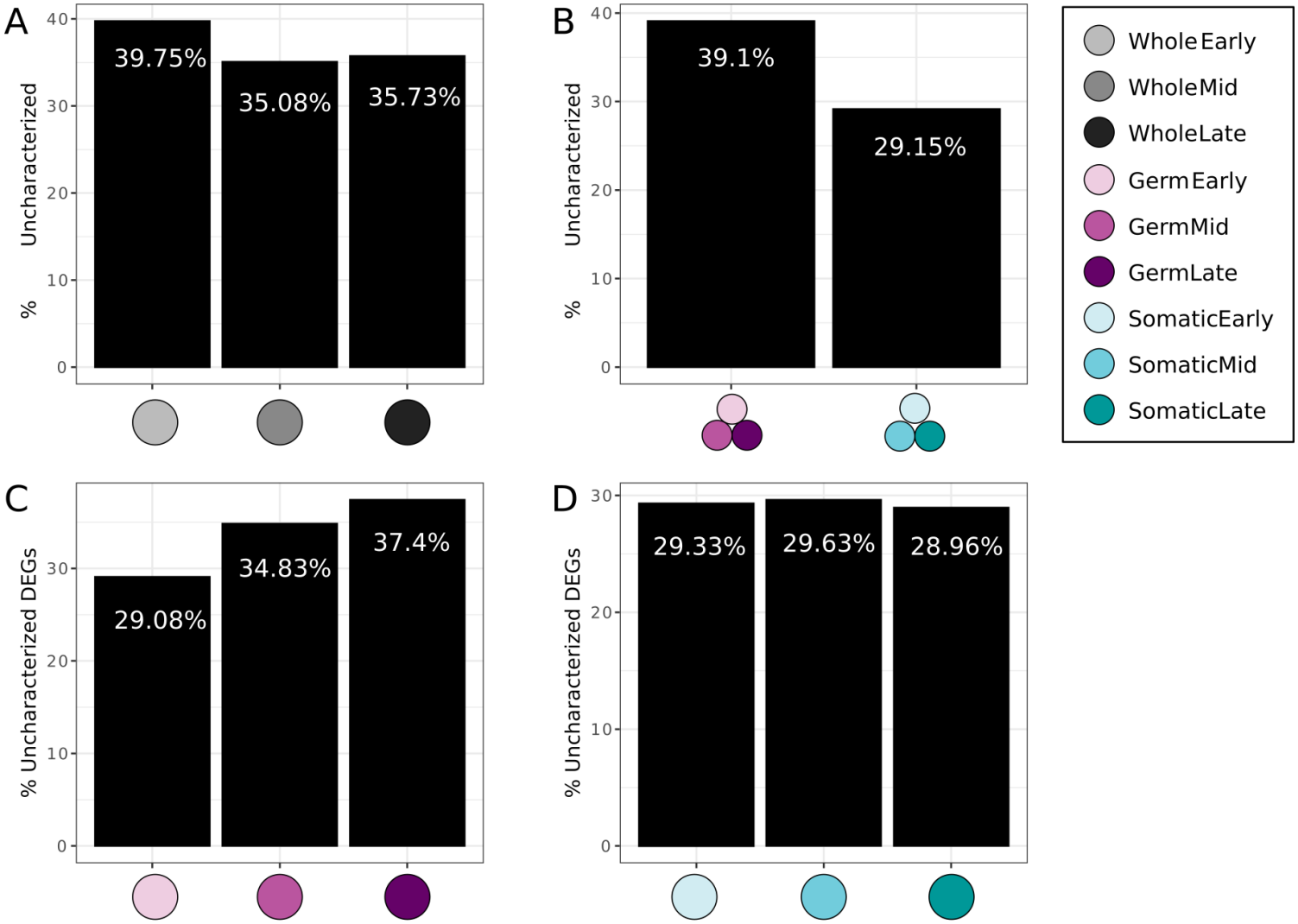
Supplementary Figure S6. Number of significantly upregulated genes in germ cells and somatic cells at A) early, B) mid, and C) late stages, and the KEGG pathways and GO-term identified as significantly enriched (BH adjusted p-value<0.05 for KEGG and BH p-value<0.01 for GO) within each set of upregulated genes. See Supplementary Table S5 for gene list.



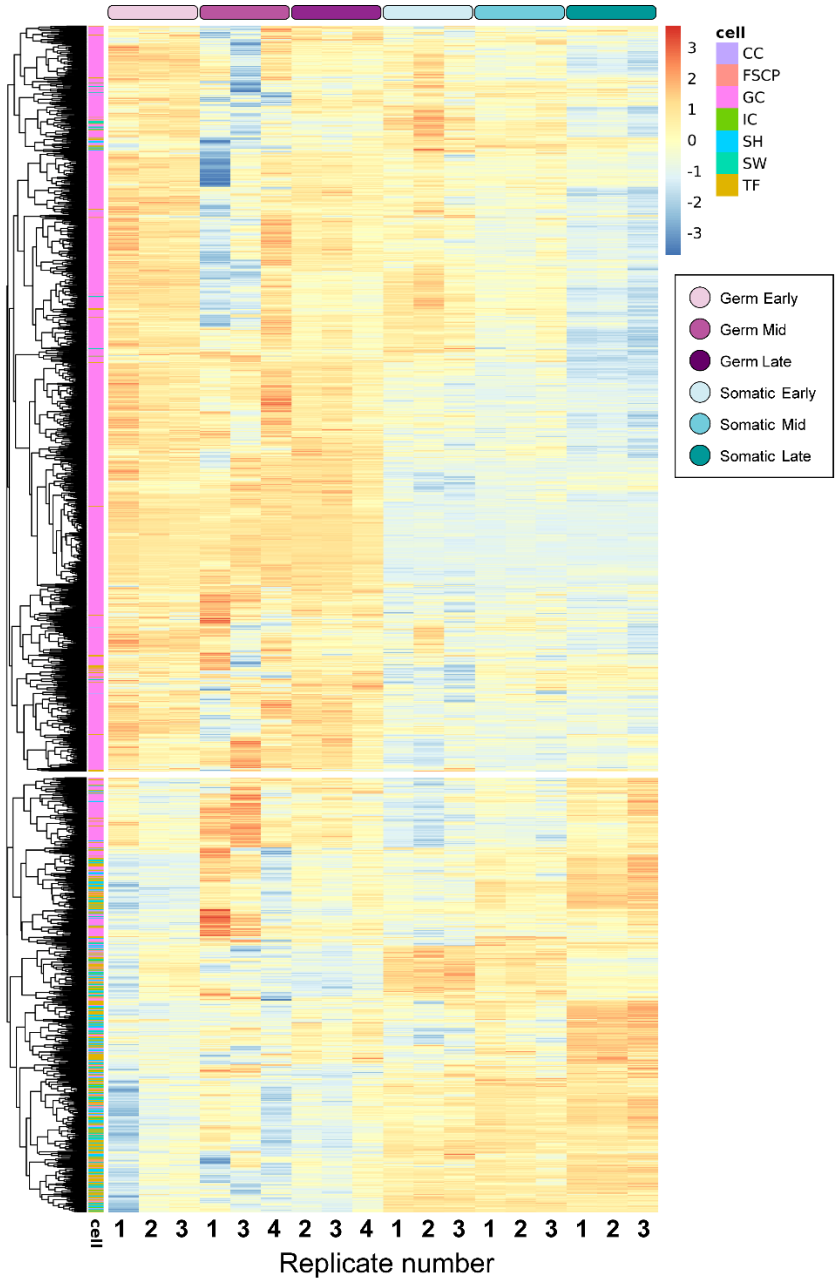
Supplementary Figure S7. Expression in VST counts of the Hippo signaling pathway core component genes according to FlyBase (FBgg0000913) in each tissue-specific RNA-seq library. Note that y axes differ slightly between plots



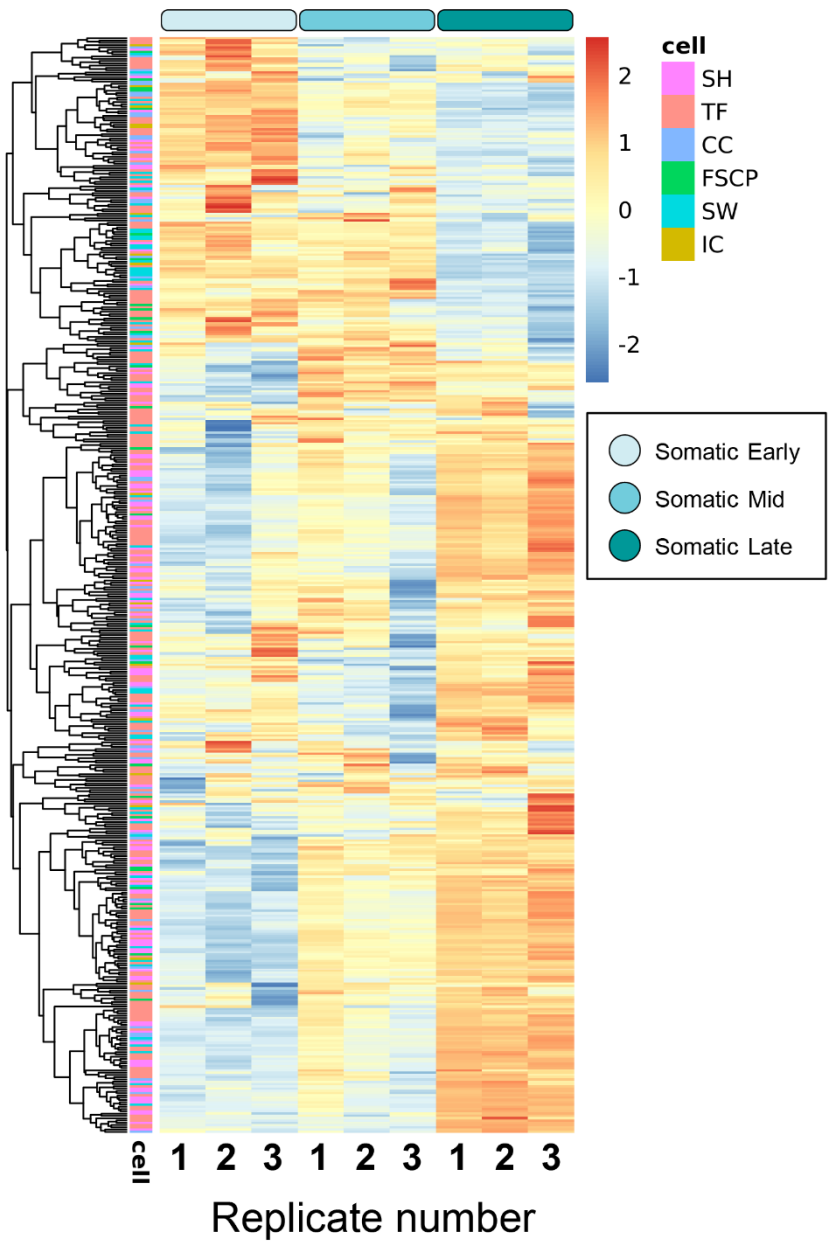
Supplementary Figure S8. Significantly enriched (BH adjusted p-value<0.05) KEGG pathways within the sets of genes significantly overexpressed (p-value<0.01) in each stage of the germ cell libraries. The circle size is proportional to the number of genes with the GO-term in the corresponding gene set. The color of the circle indicates the adjusted p-value.



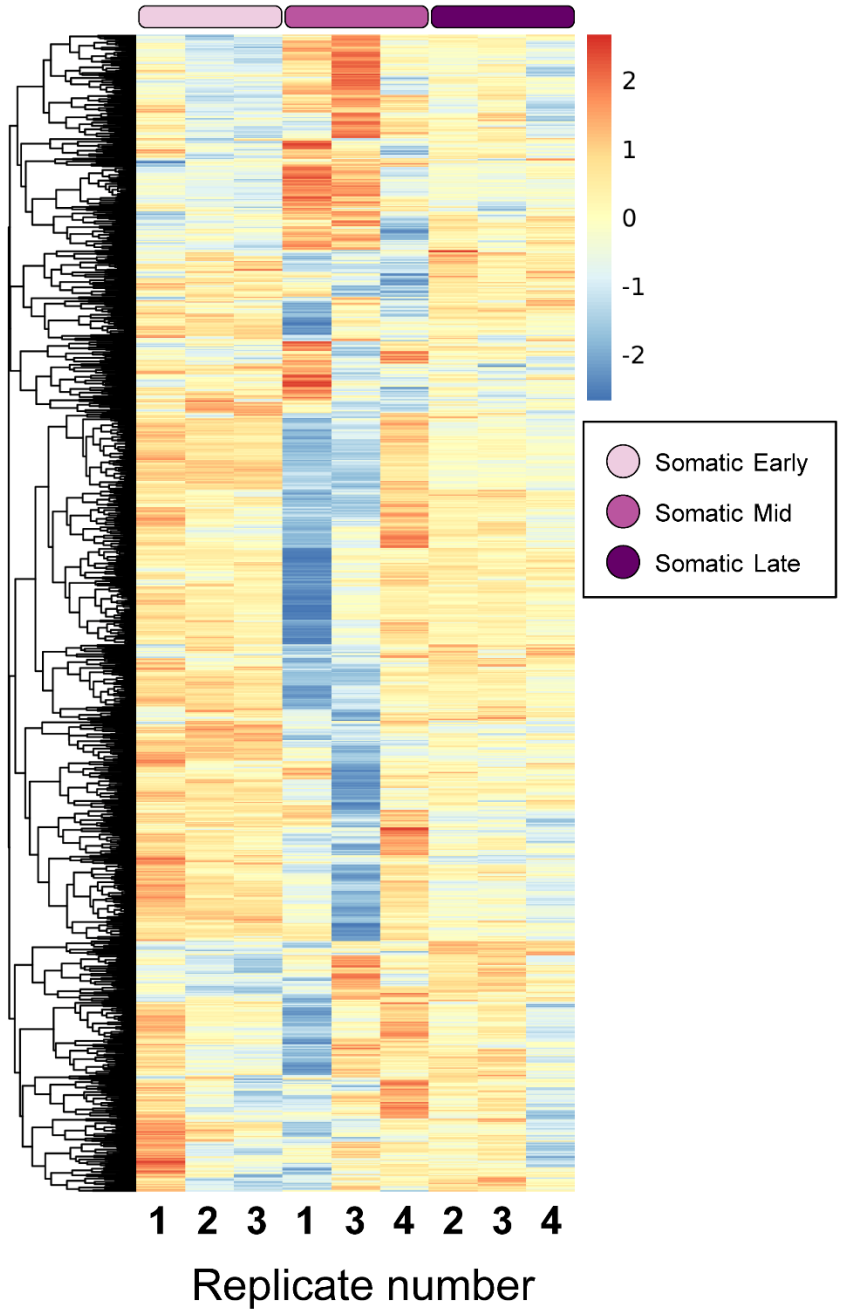
Supplementary Figure S9. Uncharacterized genes in differential expression comparisons.



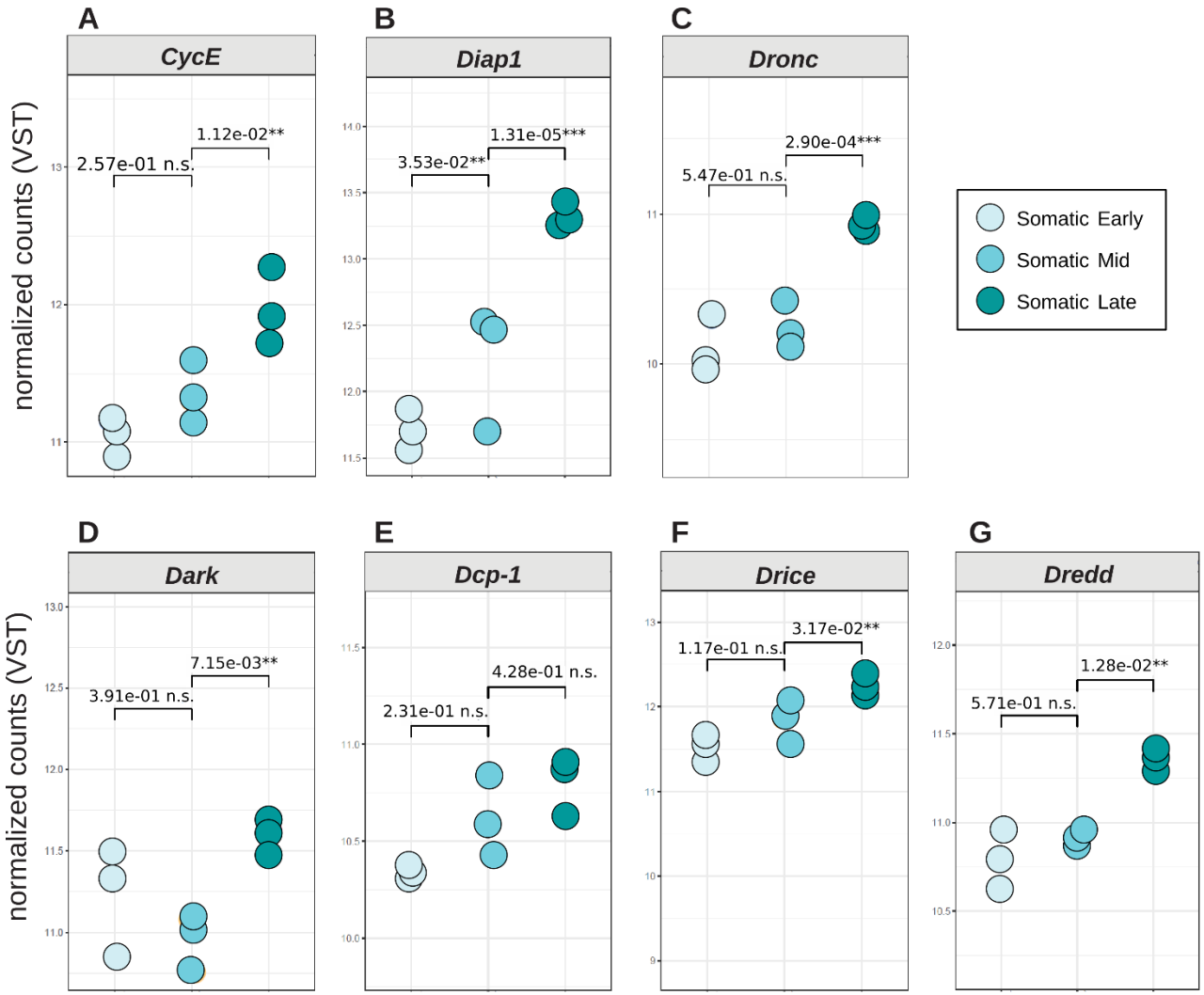
Supplementary Figure S10. Expression across our dataset of the cell markers exclusive of a single cell type obtained from SLAIDINA *et al.* (2020). The color of the first indicates the cell type of which each gene is a marker of (CC: cap cells, FSCP: follicle stem cells, GC: germ cells, IC: intermingled cells, SH: sheath cells, SW: swarm cells, TF: terminal filament). Genes are clustered based on hierarchical clustering and separated into two groups using the cuttree function which resulted in the separation of the germ cell markers from somatic markers. The expression of each gene across samples is represented as a row-wise Z-Score value of the VST-normalized counts from high(red) to low (blue).



Supplementary Figure S11. Expression across stages in somatic cell libraries of the somatic markers exclusive of a single cell type obtained from SLAIDINA *et al.* (2020). The color of the leftmost column indicates the cell type suggested by each marker gene (CC: cap cells; FSCP: follicle stem cells; GC: germ cells; IC: intermingled cells; SH: sheath cells; SW: swarm cells; TF: terminal filament). Genes are grouped based on hierarchical clustering. The expression of each gene across samples is represented as a row-wise Z-Score value of the VST-normalized counts from high (red) to low (blue).



Supplementary Figure S12. Expression across stages in germ cell libraries of the germ cell markers obtained from SLAIDINA *et al.* (2020). Genes are grouped based on hierarchical clustering. The expression of each gene across samples is represented as a row-wise Z-Score value of the VST-normalized counts from high (red) to low (blue).



Supplementary Figure S13 Dot plots of selected proliferation and apoptosis control genes differentially expressed across stages. Expression in VST counts of genes in each tissue-specific RNA-seq library.