Emergent dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of *Drosophila* testes

Amelie A. Raz[1#], Gabriela S. Vida[2#], Sarah R. Stern[3#], Sharvani Mahadevaraju[4#], Jaclyn M. Fingerhut[1#], Jennifer M. Viveiros[5#], Soumitra Pal[6#], Jasmine R. Grey[5#], Mara R. Grace[5#], Cameron W. Berry[3], Hongjie Li[7], Jasper Janssens[8], Wouter Saelens[9], Zhantao Shao[10], Chun Hun[10], Yukiko M. Yamashita[1], Teresa M. Przytycka[6], Brian Oliver[4], Julie A. Brill[11,12,13], Henry M. Krause[10,12], Erika L. Matunis[5], Helen White-Cooper[14], Stephen DiNardo[2]*, and Margaret T. Fuller[3,15]*

[1]Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology; Howard Hughes Medical Institute, Cambridge, MA 02139, USA
[2]Department of Cell and Developmental Biology, The Perelman School of Medicine and The Penn Institute for Regenerative Medicine, Philadelphia, PA 19104, USA.
[3]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305 USA.
[4]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20814, USA.
[5]Department of Cell Biology, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA.
[6]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.
[7]Huffington Center on Aging and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.
[8]JVIB Center for Brain & Disease Research, and the Department of Human Genetics, KU Leuven, Leuven, Belgium.
[9]Data Mining and Modeling for Biomedicine, VIB Center for Inflammation Research, and Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium.
[10]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada.
[11]Cell Biology Program, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada.
[12]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.
[13]Institute of Medical Science, University of Toronto, Toronto, Ontario M5S 1A8, Canada.
[14]School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3AT, UK.
[15]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305 USA.

*Correspondence to:

sdinardo@pennmedicine.upenn.edu ORC ID: 0000-0003-4131-5511

and

mtfuller@stanford.edu ORC ID: 0000-0002-3804-4987

#contributed equally

Running title: adult fly testis atlas

Key words: stem cell / lineage / spermatogenesis / snRNA-seq / scRNA-seq

Abstract

Proper differentiation of sperm from germline stem cells, essential for production of the next generation, requires dramatic changes in gene expression that drive remodeling of almost all cellular components, from chromatin to organelles to cell shape itself. Here we provide a single nucleus and single cell RNA-seq resource covering all of spermatogenesis in *Drosophila* starting from in-depth analysis of adult testis single nucleus RNA-seq (snRNA-seq) data from the Fly Cell Atlas (FCA) study (Li *et al.*, 2022). With over 44,000 nuclei and 6,000 cells analyzed, the data provide identification of rare cell types, mapping of intermediate steps in differentiation, and the potential to identify new factors impacting fertility or controlling differentiation of germline and supporting somatic cells. We justify assignment of key germline and somatic cell types using combinations of known markers, *in situ* hybridization, and analysis of extant protein traps. Comparison of single cell and single nucleus datasets proved particularly revealing of dynamic developmental transitions in germline differentiation. To complement the web-based portals for data analysis hosted by the FCA, we provide datasets compatible with commonly used software such as Seurat and Monocle. The foundation provided here will enable communities studying spermatogenesis to interrogate the datasets to identify candidate genes to test for function *in vivo*.

Introduction

Single cell RNA-seq (scRNA-seq) of developing tissues can reveal new cell types as well as previously unknown steps in the differentiation of lineages underlying tissue homeostasis and repair. In fact, high-resolution expression maps are being created for entire organisms, from *C. elegans*, planaria, and schistosomes to *Drosophila* and mouse (Cao et al., 2017, 2019; Fincher et al., 2018; Li et al., 2022; Plass et al., 2018; Schaum et al., 2018; Sebé-Pedrós et al., 2018; Siebert et al., 2019; Wendt et al., 2020), with such atlases providing a foundational reference for several important model organisms. In particular, for tissues maintained by stem cell lineages, scRNA-seq can identify the developmental trajectories that lead from dedicated tissue stem cell to terminally differentiated cell types, an important resource for understanding tissue maintenance, repair, and the origins of cancer.

The testis harbors a highly active, unipotent adult stem cell lineage that must produce sperm throughout reproductive life. Spermatogenesis relies on self-renewing germline stem cells, the progeny of which differentiate into one of the most highly specialized cell types in the body. Production of functional sperm requires intimate interactions between germ cells and somatic support cells, with defects at almost any step compromising fertility. Interest in spermatogenesis has motivated scRNA-seq analyses of testes from a variety of organisms, including mouse (Cao et al., 2021; Chen et al., 2018; Green et al., 2018; Guo et al., 2020; Law et al., 2019) and *Drosophila* (Li et al., 2022; Mahadevaraju et al., 2021; Witt et al., 2019). Notably, the testis of *Drosophila* has the highest complexity in terms of mRNAs expressed of any tissue in the fly, likely reflecting the dramatic differentiation events required (Li et al., 2022).

Many aspects of spermatogenesis are conserved from *Drosophila* to mammals. One striking difference, however, is that spermatogenesis in *Drosophila* relies on not one but two adult stem cell lineages. The co-differentiating germ cells and their closely associated somatic support cells descend from distinct stem cell populations, housed together in a well-defined niche (Fuller, 1998). Additionally, the many mutations affecting male fertility, plus powerful genetic tools for cell type specific functional analysis, have allowed identification of stage-specific regulatory factors underlying niche function in stem cell maintenance, control of proliferation, and soma-germline feedback circuits that act during co-differentiation of these two lineages. This comprehensive

98 knowledge of spermatogenesis offers a rich biological foundation for interpreting single nucleus
99 and single cell RNA-seq data.

100 Here we present an in-depth analysis of the testis subset of the Fly Cell Atlas (FCA) single
101 nucleus RNA-Seq (snRNA-seq) data. We supplement this with scRNA-seq from the same tissue,
102 together providing a foundational reference for the field. While several recent RNA-seq analyses
103 of *Drosophila* testes have been illuminating, they generally focused on particular stages (Gan et
104 al., 2010; Hof-Michel and Bökel, 2020; Lu et al., 2020; Mahadevaraju et al., 2021; Shi et al., 2020;
105 Vedelek et al., 2018; Witt et al., 2019). In contrast, the scale and comprehensive nature of the
106 FCA dataset allowed us to profile rare cell types, such as the stem cell niche, and to follow
107 spermatogenesis from early spermatogonia to late spermatids, a remarkable conversion of
108 precursors to highly elongated, specialized cells. We present supporting data for assignment of
109 key cell types, both germline and somatic, and show how progression through two distinct, yet
110 intimately interacting, stem cell-based lineages emerges from the changes in gene expression.

111 The data confirm and extend known features of the male germ line transcription program,
112 including cell type specific expression of many genes in spermatocytes, downregulation of X-
113 linked genes in later spermatocytes, and repression of most transcription in early spermatids. At
114 the same time, surprising new features emerged, including unexpected complexity in the somatic
115 support cell lineage. In addition, comparison of single nucleus with single cell sequencing data
116 significantly expanded our understanding of gene expression dynamics in spermatocytes and
117 spermatids as they mature. In particular, these data showed how dynamic changes in active
118 transcription reflected in the snRNA-seq can be obscured by the endowment of mRNAs stored in
119 the cytoplasm. This is especially clear in early haploid spermatids, which have little transcriptional
120 activity but contain many mRNAs transcribed in spermatocytes and stored to be translated later
121 to support spermatid morphogenesis. With a gene expression framework for the two testis stem
122 cell lineages now in place, mining the snRNA-seq data for changes in gene expression as one
123 cluster advances to the next should identify new sub-stage-specific markers, thereby opening the
124 way for tests of function for such newly identified genes in male germ cell differentiation.

125

126 # Results

127 ## Clustering by gene expression signature reveals progression of differentiation in
128 two stem cell lineages

129 Spermatogenesis in *Drosophila* involves obligate, intimate interactions between cells
130 differentiating in two adult stem cell-founded lineages. Male germline stem cells (GSCs) and their
131 partners, the somatic cyst stem cells (CySCs), are both physically anchored to a small cluster of
132 somatic cells termed the apical hub (Figure 1A), which provides short-range niche signals
133 important for maintenance of the two stem cell states. The interleaved arrangement of GSCs and
134 CySCs ensures that their immediate daughters are positioned to interact. Two postmitotic early
135 cyst cells enclose each gonialblast (immediate GSC daughter), forming a two-cell squamous
136 epithelium that soon seals the progeny of the gonialblast off from the rest of the testis (Fairchild
137 et al., 2015). The gonialblast initiates four rounds of spermatogonial transit amplifying divisions
138 with incomplete cytokinesis, producing 16 interconnected germ cells (Figure 1A). After the fourth
139 mitosis, the germ cells undergo premeiotic S-phase and enter an extended G2 cell cycle phase
140 termed meiotic prophase. Over the next three and a half days the 16 primary spermatocytes
141 increase 25-fold in volume, engaging in a robust transcription program in preparation for the
142 meiotic divisions and the extensive elongation and remodeling of the resulting 64 haploid
143 spermatids into mature sperm. Although they do not divide, the two somatic cyst cells co-
144 differentiate with the germ cells they enclose (Gönczy et al., 1992), eventually taking on different
145 identities as head and tail cyst cells. The head cyst cell cups the nuclear end of elongating
146 spermatid bundles and eventually inserts into the terminal epithelium at the base of the testis,
147 while the tail cyst cell elongates extensively to cover the rest of the spermatid bundle (Tokuyasu
148 et al., 1972). All these cell types, as well as somatic structural cells of the testis sheath (muscle
149 and pigment cells) and cells of the seminal vesicle are represented in the FCA testis dataset.

150 The relative similarity and differences in gene expression for 44,621 single nuclei from
151 triplicate 10X snRNA-seq runs from adult testis plus seminal vesicle (See Materials and Methods)
152 can be visualized in a Uniform Manifold Approximation and Projection (UMAP)-based
153 dimensionality reduction plot (Figure 1B). The geography of the UMAP is dominated by the

Figure 1

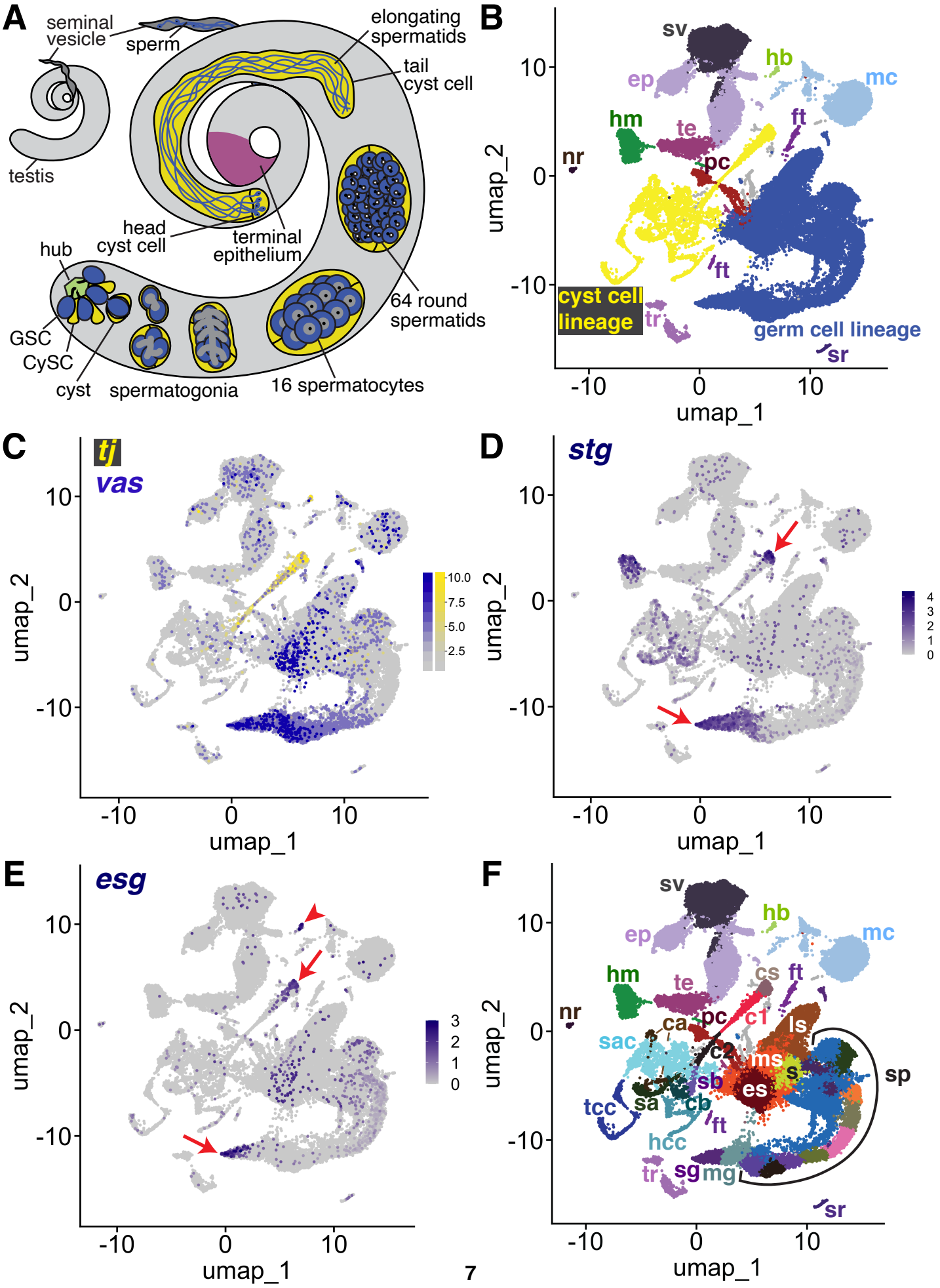
A
seminal vesicle
sperm
elongating spermatids
tail cyst cell
testis
head cyst cell
terminal epithelium
hub
64 round spermatids
GSC
CySC
cyst
spermatogonia
16 spermatocytes
B
sv
hb
mc
ep
hm
te
ft
nr
pc
ft
cyst cell lineage
tr
germ cell lineage
sr
umap_2
umap_1
C
tj
vas
10.0
7.5
5.0
2.5
D
stg
4
3
2
1
0
E
esg
3
2
1
0
F
sv
hb
mc
ep
hm
te
cs
ft
nr
ca
pc
c1
ls
sac
c2
ms
s
sp
sb
es
tcc
sa
cb
ft
hcc
tr
sg
mg
sr
umap_2
umap_1

Figure 1

The snRNA-seq landscape of the testis

A) Illustration of adult *Drosophila* testis showing hub (green), germ cell lineage (blue), cyst cell lineage (yellow), terminal epithelium (pink) and seminal vesicle (gray). **B)** UMAP of FCA snRNA-seq data from the testis plus seminal vesicle (relaxed version). Blue: germ cell lineage; Yellow: cyst cell lineage; Pink: terminal epithelial cells of testis (te); Dark gray: seminal vesicle (sv). Other cell types as listed in **F**. **(C-E**) UMAP plots of snRNA-seq data showing expression of: **C)** *traffic jam* (*tj*) (yellow) and *vasa* (*vas*) (blue), **D)** *string* (*stg*), **E)** *escargot* (*esg*). Red arrows: proliferating cells, red arrowhead: hub. **F)** UMAP (as in **B**) with Leiden 6.0 clusters of germ and cyst cell lineages labeled (sg: Spermatogonium; mg: Mid-late proliferating spermatogonia; sp: Spermatocytes; s: Spermatids; es: Early elongation-stage spermatids; ms: Early-mid elongation-stage spermatids; ls: Mid-late elongation-stage spermatids; hb: Germinal proliferation center hub; cs: Cyst stem cell, c1: Early cyst cell 1; c2: Early cyst cell 2; sa: Cyst cell with spermatocytes branch A; sb: Cyst cell with spermatocytes branch B; ca: Cyst cell branch a; cb: Cyst cell branch b; sac: Elongating spermatid-associated cyst cell; hcc: Head cyst cell; tcc: Tail cyst cell; te: Terminal epithelial cells of testis; sv: Seminal vesicle; ep: Male gonad associated epithelium; sr: Secretory cells of the male reproductive tract; mc: Muscle cell; hm: Hemocyte; nr: Neuron; pc: Pigment cell;, tr: Trachea; ft: Fat body).

154 dynamic sequences of differentiating states in the germline (blue) and somatic cyst cell (yellow)
155 lineages. Each lineage manifests as an emergent trajectory of nuclei with continuously
156 progressing gene expression profiles, unlike the discrete clusters characteristic of most terminally
157 differentiated cell types. Despite their physical proximity and cooperation *in vivo*, the germ line
158 and cyst cell lineages mapped to largely non-overlapping formations in gene expression space
159 represented in the UMAP, consistent with their different embryological origin, cell biology, and
160 known roles.

161 From the perspective in Figure 1B, the spatial arrangement of nuclei in the UMAP
162 whimsically resembles a hammerhead shark (blue - germ line) playing a saxophone (yellow - cyst
163 cell lineage) watched over by a mermaid (several somatic epithelial-based structural elements,
164 including the seminal vesicle (sv, dark purple) and terminal epithelial cells at the testis base (te,
165 pink)). One notable cluster located near the mermaid head is the hub (hb, light green), the niche
166 that supports the two stem cell lineages. Other clusters on the UMAP contain differentiated cell
167 types that contribute to organ structure, including muscle (mc) and pigment cells (pc) of the testis
168 sheath (Figure 1B). Additionally, sample dissection carried over small numbers of non-testis cells,
169 including tracheal (tr) and fat body (ft) cells, hemocytes (hm), neurons (nr), and male reproductive
170 tract secretory cells (sr).

171 Identity of key clusters was assigned based on expression of known markers from the
172 literature (citations for all published markers employed given in Table 1). Expression of *vasa (vas)*
173 identified early germ line nuclei while expression of *traffic jam (tj)* identified nuclei from early
174 stages in the somatic cyst cell lineage (Figure 1C). Expression of the cdc25 phosphatase *string*
175 *(stg)*, required for the G2/M transition in mitotic cells (Alphey et al., 1992; Edgar and O'Farrell,
176 1990), and *escargot (esg)*, a gene expressed in diploid proliferative cells (Fuse et al., 1994),
177 marked CySCs in the cyst cell lineage and proliferating GSCs and spermatogonia in the germ line
178 lineage (Figure 1D,E). Expression of *esg* also marked the hub, as expected from prior studies
179 (Voog et al., 2014) (Figure 1E). Together, these markers established that the germ line lineage
180 begins at the tail end of the "shark" with germ line stem cells (GSCs) and proliferating
181 spermatogonia at the tapered point. The somatic cyst cell lineage begins at the mouthpiece of

TABLE 1:	List of genes used as markers in identifying key clusters			
Gene_symbol	**Gene_name**	**FBgn**	**Reference**	**DOI**
aly	always early	FBgn0004372	White-Cooper et al., 2000	https://doi.org/10.1242/dev.127.24.5463
aub	aubergine	FBgn0000146	Nishida et al., 2007	DOI: 10.1261/rna.744307
bam	bag of marbles	FBgn0000158	Schulz et al., 2004	DOI: 10.1534/genetics.103.023184
CadN	Cadherin-N	FBgn0015609	Boyle et al., 2007	DOI: 10.1016/j.stem.2007.08.002
can	cannonball	FBgn0011569	Hiller et al., 2001	DOI: 10.1101/gad.869101
cher	cheerio	FBgn0014141	Tanentzapf et al., 2007	DOI: 10.1038/ncb1660
CycB	Cyclin B	FBgn0000405	White-Cooper et al., 1998	DOI: 10.1242/dev.125.1.125
Dic61B	Dynein intermediate chain at 61B	FBgn0263988	Lu et al., 2020	DOI: 10.1101/gad.335331.119
dlg1	discs large 1	FBgn0001624	Papagianoulli and Mechler, 2009	DOI: 10.1038/cr.2009.71
esg	escargot	FBgn0287768	Kiger et al., 2000	10.1038/35037606
eya	eyes absent	FBgn0000320	Fabrizio et al., 2003	10.1016/s0012-1606(03)00127-1
f-cup	flyers-cup	FBgn0028487	Barreau et al., 2008	DOI: 10.1242/dev.021949
Fas3	Fasciclin III	FBgn0000636	Brower et al., 1981	https://doi.org/10.1242/dev.63.1.233
fzo	fuzzy onions	FBgn0011596	Hwa et al., 2002	DOI: 10.1016/s0925-4773(02)00141-7
hh	hedgehog	FBgn0004644	Michel et al., 2012	10.1242/dev.075242
Hml	Hemolectin	FBgn0029167	Li et al., 2022	DOI: 10.1126/science.abk2432
kl-2	male fertility factor kl2	FBgn0001313	Carvalho et al, 2000	10.1073/pnas.230438397
kl-3	male fertility factor kl3	FBgn0267432	Carvalho et al, 2001	10.1073/pnas.230438398
kl-3	male fertility factor kl3	FBgn0267432	Fingerhut et al., 2019	10.1371/journal.pgen.1008028
kl-5	male fertility factor kl5	FBgn0267433	Gepner and Hays, 1993	10.1073/pnas.90.23.11132
kl-5	male fertility factor kl5	FBgn0267433	Fingerhut et al., 2019	10.1371/journal.pgen.1008028
kmg	kumgang	FBgn0032473	Kim et al., 2017	DOI: 10.1126/science.aal3096
Mst77F	Male-specific transcript 77F	FBgn0086915	Barckmann et al., 2013	https://doi.org/10.1016/j.ydbio.2013.02.018
Mst84Db	Male-specific RNA 84Db	FBgn0004173	Kuhn et al., 1991	DOI: 10.1016/0925-4773(91)90064-d
Mst84Dc	Male-specific RNA 84Dc	FBgn0004174	Kuhn et al., 1991	DOI: 10.1016/0925-4773(91)90064-d
Mst87F	Male-specific RNA 87F	FBgn0002862	Kuhn et al., 1991	DOI: 10.1016/0925-4773(91)90064-d
MtnA	Metallothionein A	FBgn0002868	Zhao et al., 2010	doi: 10.1093/nar/gkp1006
Nep5	Neprilysin 5	FBgn0039478	Sitnik et al., 2014	DOI: 10.1534/genetics.113.160945
p53	p53	FBgn0039044	Monk et al., 2012	DOI: 10.1007/s00441-012-1479-4
p-cup	presidents-cup	FBgn0030840	Barreau et al., 2008	DOI: 10.1242/dev.021949
piwi	P-element induced wimpy testis	FBgn0004872	Gonzalez et al., 2015	DOI: 10.1016/j.celrep.2015.06.004
Rbp4	RNA-binding protein 4	FBgn0010258	Baker et al., 2015	https://doi.org/10.1242/dev.122341
sa	spermatocyte arrest	FBgn0002842	Hiller et al., 2004	DOI: 10.1242/dev.01314
shg	shotgun	FBgn0003391	Voog et al., 2008	DOI: 10.1038/nature07173
so	sine oculis	FBgn0003460	Fabrizio et al., 2003	DOI: 10.1016/s0012-1606(03)00127-1
soti	scotti	FBgn0038225	Barreau et al., 2008	DOI: 10.1242/dev.021949
stg	string	FBgn0003525	Alphey et al., 1992	https://doi.org/10.1016/0092-8674(92)90616-K
Syt1	synaptotagmin 1	FBgn0004242	Li et al., 2022	DOI: 10.1126/science.abk2432
tj	traffic jam	FBgn0000964	Li et al., 2003	DOI: 10.1038/ncb1058
tomboy20	tomboy20	FBgn0037828	Hwa et al., 2004	https://doi.org/10.1016/j.febslet.2004.07.025
upd1	unpaired 1	FBgn0004956	Tulina and Matunis 2001	DOI: 10.1126/science.1066700
vas	vasa	FBgn0283442	Hay et al., 1988	DOI: 10.1016/0092-8674(88)90216-4
wa-cup	walker cup	FBgn0037502	Barreau et al., 2008	DOI: 10.1242/dev.021949
zpg	zero population growth	FBgn0024177	Tazuke et al., 2002	DOI: 10.1242/dev.129.10.2529

182 the “saxophone” at the UMAP center, with early cyst cell nuclei extending down and leftward in a
183 thin line. In addition, analysis by fluorescence *in situ* hybridization (FISH), and the average
184 number of unique transcripts (Unique Molecular Identifier - UMI) expressed helped assign identity.
185 For example, spermatocytes are highly transcriptionally active, whereas *Drosophila* early
186 spermatids are nearly quiescent. While clustering was carried out using the Leiden algorithm at
187 increasing levels of resolution, we settled on Leiden 6.0 as providing optimal granularity along
188 both somatic and germline differentiation trajectories. We assigned 43 clusters as germline and
189 22 clusters as likely cyst cell lineage, with many inferred from the UMAP geography as
190 representing putative intermediate cell types in the respective lineages (Figures 1F, 2A, 6A, and
191 S1A).

192 **Progression of differentiation in the male germ line stem cell lineage**

193 Figure 2A shows the UMAP for the germ line stem cell lineage with Leiden 6.0 clusters
194 labeled. Expression of the germ cell-specific gap junction gene *zero population growth* (*zpg),*
195 required for survival of early spermatogonia (Tazuke et al., 2002) (Figure 2D, L), along with *vasa,*
196 *stg* and *esg* (Figure 1C, D and E), further established nuclei at the pointed tip of the shark tail
197 (clusters 25 and 22) as GSCs and spermatogonia. *In vivo*, GSCs are distinguished from
198 gonialblasts and transit amplifying spermatogonia cytologically, by attachment to the apical hub
199 and cell biological characteristics such as oriented centrosomes and spindles, and functionally,
200 by lineage analysis. However, mRNA markers restricted to GSCs have not yet been identified,
201 preventing us from determining what percent of these early nuclei are GSCs. Many nuclei in
202 cluster 22 express *bag-of-marbles (bam)* (Figure 2L) but lack known spermatocyte markers,
203 suggesting that these nuclei represent mid-to-late spermatogonia or germ cells undergoing
204 premeiotic S phase. Moving rightward, several known early spermatocyte markers such as
205 *kumgang* (*kmg)* and *RNA-binding protein 4* (*Rbp4*) began to be expressed (clusters 5, 78, and 40)
206 (Figure 2F-G, L). Transcripts from *aubergine* (*aub),* a piRNA binding protein, were detected in
207 the spermatogonial region (clusters 25, 22) and overlapping with early spermatocyte markers
208 (clusters 5, 78, 40 and 41) (Figure 2E and L). Fluorescent *in situ* hybridization (FISH) confirmed
209 *aub* transcripts present in GSCs around the hub, spermatogonia, and extending into early

Figure 2

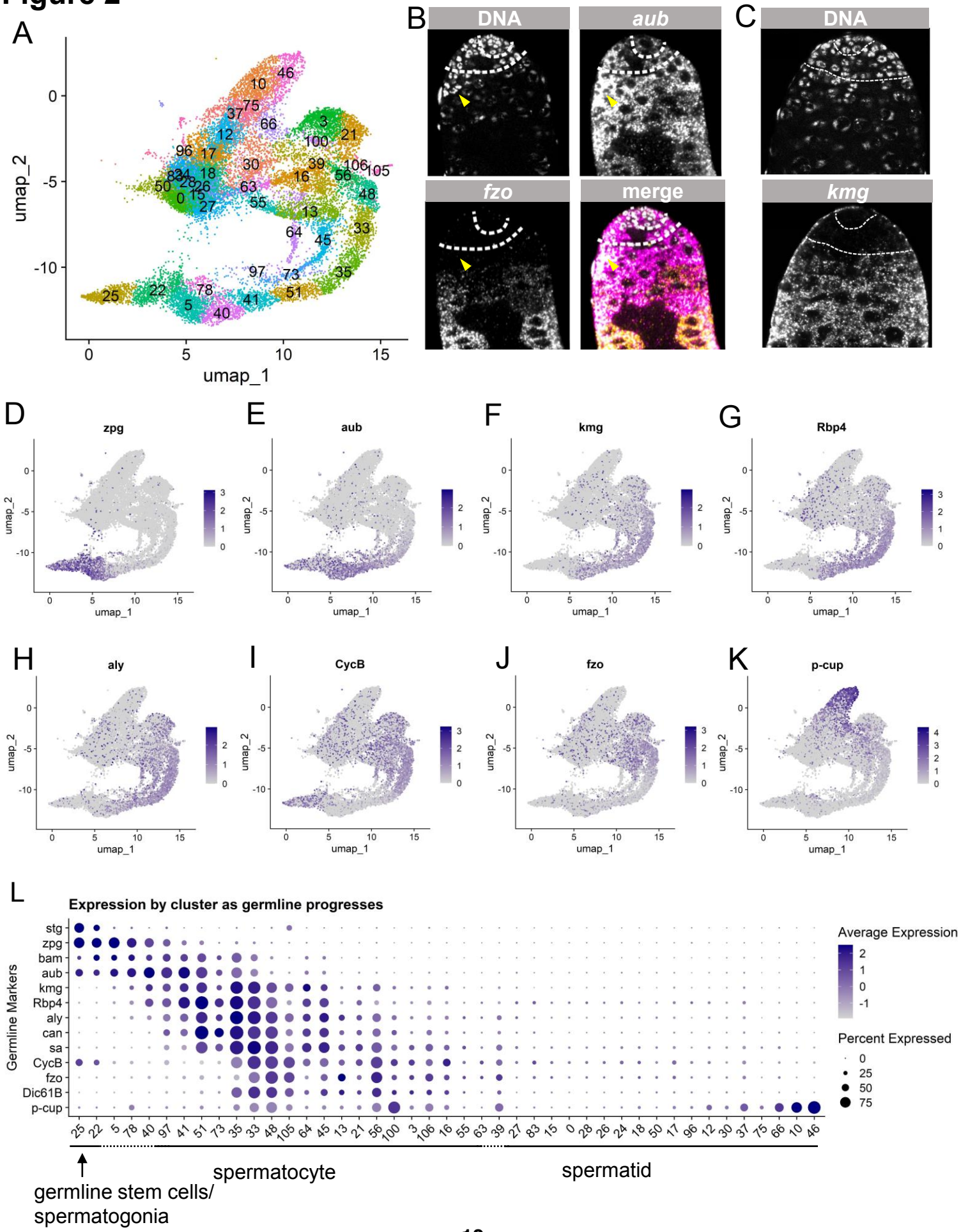
A
umap_2
umap_1
B
DNA
aub
fzo
merge
C
DNA
kmg
D
zpg
E
aub
F
kmg
G
Rbp4
H
aly
I
CycB
J
fzo
K
p-cup
L
Expression by cluster as germline progresses
Germline Markers
stg
zpg
bam
aub
kmg
Rbp4
aly
can
sa
CycB
fzo
Dic61B
p-cup
Average Expression
Percent Expressed
spermatocyte
spermatid
germline stem cells/
spermatogonia

Figure 2 with one supplement

Characteristics of the germline lineage

A) Germline portion of the UMAP generated by Seurat from clustering of the full testis plus seminal vesicle dataset at Leiden 6.0 resolution. **B-C)** Apical tips of testes showing localized expression of **B)** *aub* (magenta) and *fzo* (yellow) mRNA and **C)** *kmg* mRNA visualized by *in situ* hybridization. Apical-most dotted line demarcates germ line stem cells (GSCs) around the hub from spermatogonia. Lower dotted line demarcates spermatogonia from young spermatocytes. Yellow arrowhead in **B**: early spermatocytes. **D-K)** Feature plots generated by Seurat showing expression levels of *zpg, aub, kmg, rbp4, aly, CycB, fzo,* and *p-cup* in the germline UMAP. Navy blue gradient bars: relative expression level for the indicated gene. **L)** Dot plot generated by Seurat showing expression levels of selected germline markers by cluster as nuclei progress from spermatogonia to spermatid. Color intensity: level of expression of the indicated gene averaged over all the nuclei in a given cluster relative to the level for that gene in other clusters. Size of dots: percent of nuclei in each cluster in which expression of the gene was detected. (see also FIgure 2 - figure supplement 1).

210 spermatocyte cysts, with their characteristic larger nuclei (Figure 2B). FISH also confirmed
211 expression of *kmg* mRNA starting in early spermatocytes (Figure 2C), with early spermatocytes
212 showing both *aub* and *kmg* transcripts, consistent with the snRNA-seq data.

213 Progressively maturing spermatocytes along the bottom right of the germline UMAP
214 expressed later markers, including mRNAs for the spermatocyte-specific tMAC subunit *always*
215 *early* (*aly)* and the testis-specific TAFs (tTAFs) *spermatocyte arrest (sa)* and *cannonball* (*can*)
216 (clusters 41, 51, 35, and onward; Figure 2H,L). Expression of *fuzzy onions* (*fzo)* and *Dynein*
217 *intermediate chain 61B* (*Dic61B*) was detected later, as the germ cell clusters curved upward
218 (clusters 33, 48,105, 45, 13, 56; Figure 2J,L), consistent with the dependence of *fzo* and *Dic61B*
219 transcription on *aly* (Hwa et al., 2002; Lu et al., 2020). Correlating *in vivo* morphology with gene
220 expression space (visualized in the UMAP), *fzo* transcripts were not detected by FISH in the
221 young spermatocytes near the spermatogonial region but were strongly detected in more mature
222 spermatocytes further away from the testis apical tip (Figure 2B). The G2/M cell cycle regulator
223 *CyclinB* (*CycB*) is transcribed from one promoter in mitotic spermatogonia, silenced, then re-
224 expressed from an alternate promoter in later spermatocytes, dependent on *aly* function (Lu et
225 al., 2020; White-Cooper et al., 1998). These two distinct stages of *CycB* transcript expression are
226 clearly visible in the snRNA-seq data (Figure 2I,L). Maturing spermatocytes, marked by
227 expression of Y-linked genes (Figure 3D), lie toward the top of the upward curve where the tail
228 meets the torso of the shark. The progression of germ cell differentiation continues with early
229 stage spermatids along the upper torso and head of the shark (marked by low UMI - see below).
230 Mid-to-late elongation stage spermatids, marked by expression of *p-cup* mRNA, lie in the blunt
231 projection toward the upper right of the UMAP (Figure 2K,L; clusters 66, 10, 46).

232 The order of clusters in expression space reflects differentiation in the lineage, as
233 indicated by plotting the expression of known germline markers in each UMAP cluster (Figure 2L).
234 Notably, using the published marker genes scored here, sequential cluster identities (e.g., 25, 22,
235 5, 78, 40) were not each delimited by unique marker genes. Instead, graded expression of the
236 markers examined extended across boundaries between clusters. This was also observed in a
237 UMAP with just nine clusters created at lower resolution (Figure S1B,C).

238 The geography of the UMAP is reminiscent of the spatio-temporal organization in the
239 testis itself, with stages laid out from GSCs to transit amplifying spermatogonia, then young, mid,
240 and late-stage spermatocytes. However, as the UMAP displays changes in gene expression
241 rather than physical space, some surprisingly long stretches of the UMAP represent what are
242 known to be relatively short periods in developmental time. For example, a large stretch along
243 the bottom of the UMAP (clusters 5, 78, 40, 41, 51) represents young spermatocytes, previously
244 thought of as a single, short developmental stage. This is underscored by the long gap in
245 detection of *CycB* mRNA in these clusters (Figure 2I, L). In contrast, *in situ* hybridization showed
246 only a relatively narrow region near the boundary between spermatogonia and spermatocytes
247 devoid of *CycB* mRNA (White-Cooper et al., 1998). The territory of the UMAP containing clusters
248 5, 78, 40, 41, 97, 73, and 51 may be stretched out in gene expression space because the early
249 spermatocyte stages are a time of extensive, rapid, and dynamic changes in gene expression:
250 many genes are being dramatically upregulated as the spermatocyte expression program
251 initiates, while a number of genes transcribed in spermatogonia are being downregulated (Figure
252 2B-H,L; Figure 3A; orange, tan and green cells in Figure 4A,C,F,G - see also Shi et al., 2020).
253 Indeed, young spermatocytes are represented by fewer nuclei (2462; clusters 5, 78, 40, 41, 51) then
254 more mature spermatocytes (~4100 nuclei; clusters 35, 33, 64, 45, 48, 105, 56, 13, 16). In addition, as
255 early spermatocytes are still relatively small, they will occupy less physical space than more
256 mature spermatocytes.

257 **The spermatocyte transcription program**

258 The spermatocyte period features onset of dramatic transcriptional changes. Many genes
259 expressed in spermatocytes are transcribed in few or no other known cell types, including the
260 markers *kmg, Rbp4, fzo, can, sa* (see references in Table 1). This robust onset of cell type-
261 specific transcription appears as an increase in the number of different genes detected per
262 nucleus (Figure 3A,C), leading to a substantial increase in transcriptome complexity. Coincident
263 with this was a large increase in the number of unique molecular identifiers (UMI) scored, peaking
264 in mid-to-late spermatocyte nuclei (clusters 35 and 33), with average UMI per cluster increasing
265 from <5000 to >30,000 UMI per nucleus as spermatogonia differentiated to late spermatocytes

Figure 3

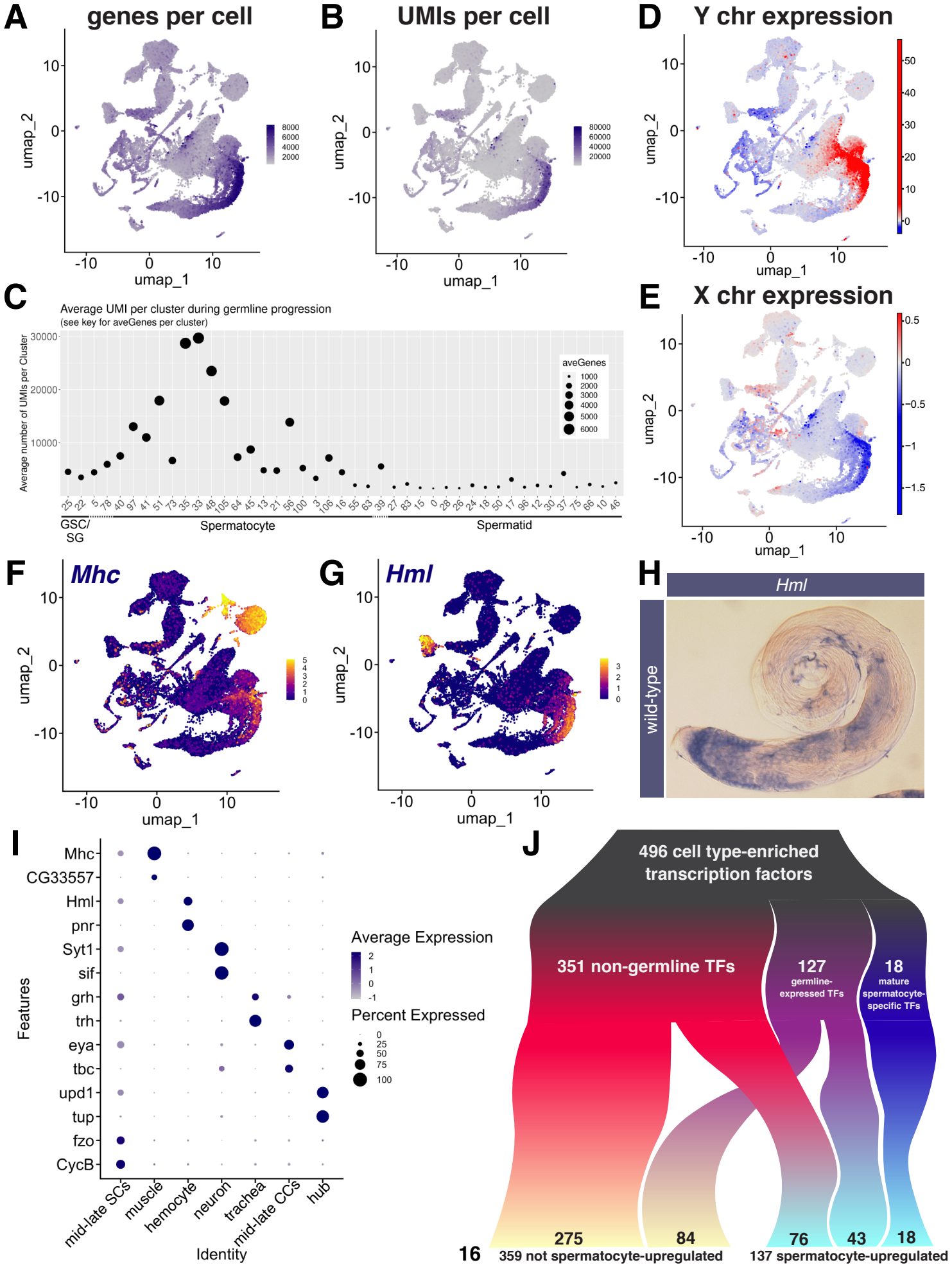
A
genes per cell
B
UMIs per cell
D
Y chr expression
C
Average UMI per cluster during germline progression
(see key for aveGenes per cluster)
Average number of UMIs per Cluster
aveGenes
GSC/
SG
Spermatocyte
Spermatid
E
X chr expression
F
Mhc
G
Hml
H
Hml
wild-type
I
Features
Mhc
CG33557
Hml
pnr
Syt1
sif
grh
trh
eya
tbc
upd1
tup
fzo
CycB
mid-late SCs
muscle
hemocyte
neuron
trachea
mid-late CCs
hub
Identity
Average Expression
Percent Expressed
J
496 cell type-enriched transcription factors
351 non-germline TFs
127 germline-expressed TFs
18 mature spermatocyte-specific TFs
275
84
76
43
18
359 not spermatocyte-upregulated
137 spermatocyte-upregulated

Figure 3 with one supplement

Features of the spermatocyte transcription program

(A,B) UMAPs of snRNA-seq data showing: **A)** number of genes detected as expressed and **B)** number of unique molecular identifiers (UMIs) detected per nucleus. **(C)** Plot of average number of genes expressed (dot size) and UMIs detected per nucleus per germline-annotated Leiden 6.0 cluster, ordered by progression of germ line differentiation. **(D,E)** UMAPs of snRNA-seq data showing the average expression of: **D)** transcripts on the Y chromosome or **E)** transcripts on the X chromosome relative to an expression-matched control set (gene sets with binned expression matching transcript lists). (Note the dramatically different scales in D vs E.) UMAPs taken directly from ASAP (Li *et al.*, 2022, Gardeux *et al.*, 2017). **(F,G)** UMAP plots of snRNA-seq raw counts (log-transformed) showing expression of: **F)** *Myosin heavy chain* (*Mhc*) in muscle and late spermatocytes and **G)** *Hemolectin* (*Hml*) in hemocytes and late spermatocytes. Yellow: relative expression high. **H)** Testis hybridized *in situ* with biotinylated antisense RNA probe to *Hml,* showing expression (blue) in spermatocytes. **I)** Dot Plot showing expression of pairs of tissue-specific markers across cell types. Average expression of each gene in a given cell type denoted by color intensity. Percent of cells of the given cell type that express each gene denoted by dot size. **J)** Flow chart showing whether 496 predicted transcription factors identified as relatively cell type specific in the FCA study (Li *et al.*, 2022) (top of chart) were detected as upregulated in spermatocytes in our analysis of testis plus seminal vesicle snRNA-seq (bottom of chart; see Figure 3 - source data 1, and Methods). (see also FIgure 3 - figure supplement 1).

266 (Figure 3B,C). The FCA paper noted that testis, heart, fat body, Malpighian tubules, and male
267 reproductive glands had relatively high RNA levels and number of genes expressed compared to
268 other tissues (Li et al., 2022). Reanalysis showed that mid-to-late spermatocyte nuclei exhibited
269 the highest complexity of all, with average expressed gene (6,000 compared to 2,000) and UMI
270 (30,000 compared to <20,000) numbers higher than for any cluster mapped in heart, Malpighian
271 tubules, or male reproductive glands (Figure 3 - figure supplement 1A; the high fat body signal
272 was due to contaminating spermatocytes). High transcriptome complexity has also been noted
273 in mammalian spermatocytes (Soumillon et al., 2013).

274 After peaking in clusters 35 and 33, UMI values per nucleus decreased through clusters
275 48, 105, 56, 106, 21, where the "tail" meets the "torso" of the shark, consistent with the observed
276 lower expression of spermatocyte marker genes (Figure 2L). In the shark upper torso and head,
277 many clusters had very low UMI (Figure 3C), making developmental order difficult to assign. This
278 is reflected in the UMAP shape, with clusters grouped rather than extended along a string as in
279 early germ cell stages. We surmise these nuclei represent early spermatids, as classic studies
280 showed that transcription falls dramatically from shortly before onset of the meiotic divisions, with
281 no bulk incorporation of radioactive uridine detected in haploid round and early elongating
282 spermatid nuclei (Gould-Somero and Holland, 1974; Olivieri and Olivieri, 1965).

283 Spermatocytes showed sex chromosome specific trends in gene expression changes.
284 Overall, Y-linked transcripts were strongly upregulated in spermatocytes (Figure 3D), primarily
285 driven by the robust expression of 8 of the 12 single copy genes. For example, transcription of
286 the Y-linked fertility factors *kl-3* and *kl-5*, which encode flagellar dyneins expressed only in male
287 germ cells, was massively upregulated (125 and 275 fold, respectively with similarly large
288 differences in absolute expression). As 10X sequencing utilizes oligo(dT) primers, the late
289 appearance of reads from the Y-linked fertility factors in spermatocytes may reflect the very long
290 time required to complete synthesis of the mature transcripts, which have extremely large introns
291 (Fingerhut et al., 2019). For X linked genes, analysis of the snRNA-seq data showed a similar
292 level of expression relative to a control set of genes from all chromosomes in spermatogonia and
293 early spermatocytes. However mid-to-late spermatocytes featured a transition to reduced

294 expression of X-linked genes relative to the control set (Figure 3E), consistent with the roughly 2
295 fold lower expression of X linked genes compared to generally expressed autosomal genes
296 observed previously (Mahadevaraju et al., 2021).

297 One surprise that emerged from the UMAP geography was that later stage spermatocytes
298 split into three parallel streams, all expressing spermatocyte specific markers. (Figure 2A).
299 Strikingly, nuclei in the leftmost and middle streams (clusters 64 and 45, respectively) had
300 considerably lower UMI count than in the robust mainstream (cluster 35; Figure 3B,C). The cause
301 underlying such different UMI levels among late spermatocytes is not known, but could suggest
302 a stochastic component to meiotic chromosome condensation and the attendant chromosome-wide
303 downregulation in gene expression.

304 A second notable feature was the expression in mid-to-late spermatocytes of many
305 markers classically associated with other cell types. Notably, markers for muscle *(Myosin heavy*
306 *chain* - *Mhc*), hemocytes (*Hemolectin* - *Hml,*) neurons (*Synaptotagmin* - *Syt1*), and epithelial cells
307 (*grainy head* - *grh*) selected as identifiers of these cell types in the FCA study of adult *Drosophila*
308 tissues (Li et al., 2022), were upregulated in late spermatocytes (Figure 3F,G,I). This was
309 confirmed by in situ for Hml (Figure 3H - figure supplement 1D,E,H). Several mRNAs normally
310 thought of as markers of somatic cells, *eyes absent* (*eya*) and *unpaired 1* (*upd*) for example, were
311 detected as upregulated in mid-to-late spermatocytes (Figure 3 - figure supplement 1F,G).
312 Expression of the RNAs was usually lower in spermatocytes than in the "marker" tissue, and not
313 all genes characteristically expressed in marker tissue were detected as expressed in
314 spermatocytes (Figure 3I, see also *tj* (Figure 1B) compared to *eya* (Figure 3 - figure supplement
315 1F). Similar upregulation of *Mhc*, *Hml*, *grh,* and *Syt1* in spermatocytes was independently
316 observed in a single cell testis dataset to be introduced below (Figure 3 - figure supplement 1I),
317 so is not likely to be an artifact of isolation of nuclei. The cause of this seemingly promiscuous
318 expression of certain genes, whether the encoded proteins accumulate, and what role, if any,
319 these genes may have in the biology of spermatocytes remain to be investigated.

320 Although spermatocytes feature an overall increase in UMIs per cell, not all genes are
321 promiscuously upregulated in spermatocytes. For example, across all adult fly tissues, the FCA

322 project identified 496 transcription factors predicted to have a high tissue specificity score (Li et
323 al., 2022). That analysis predicted 351 of these as expressed in certain somatic cell types but
324 not in germ line (Figure 3J, top). Our analysis of the testis plus seminal vesicle portion of the FCA
325 data showed that only about a fifth (76) of these 'somatic' factors were upregulated in
326 spermatocytes compared to spermatogonia, with most (275) exhibiting no upregulation (Figure
327 3J bottom; Figure 3 - source data 1). This again suggests that there is some specificity to the
328 'promiscuous expression' in spermatocytes.

329 **sn vs. scRNA-seq: dynamics of active transcription vs. stored RNAs**

330 Although largely transcriptionally silent, early spermatids carry numerous cytoplasmic
331 transcripts, many of which are recruited to be translated for temporal control of protein expression
332 during spermatid morphogenesis (Schäfer et al., 1990, 1995). As single-nucleus sequencing
333 detects recently transcribed or nuclear resident transcripts, comparison of snRNA-seq with *single*
334 *cell* RNA-seq (scRNA-seq) data should identify genes no longer actively transcribed at a particular
335 stage but represented by mRNAs retained/perduring from earlier transcription. Single cell RNA-
336 seq of testes yielded data for 6438 germ cells after quality control steps (Materials and Methods).
337 Indeed, comparison of snRNA-seq and scRNA-seq allowed mapping of dynamic changes in
338 transcriptional activity alongside identification of an extensive array of post-transcriptionally
339 retained (perduring) transcripts stored for later use.

340 The UMAP geography for both the snRNA-seq and scRNA-seq datasets showed progression
341 from spermatogonia to spermatids, with germline differentiation classes present in sequential
342 order (Figure 4A-D). In the scRNA-seq UMAP, as for snRNA-seq, expression of *zpg* marked a
343 small number of spermatogonia located at the bottom tip, *aub* marked those same plus additional
344 cells, presumably early spermatocytes, *fzo* marked differentiating spermatocytes, and *presidents-*
345 *cup* (*p-cup*) marked later elongating spermatids in an arm extending from the top of the UMAP
346 (Figure 4E). Corroborating these expression patterns, FISH to whole mount testes clearly showed
347 *aub* expression in spermatogonia and early spermatocytes, and *fzo* expression beginning in
348 spermatocytes (Figure 4F). Notably, *fzo* mRNA was abruptly downregulated in early round
349 spermatids soon after the second meiotic division (Figure 4F, arrowhead).

Figure 4.

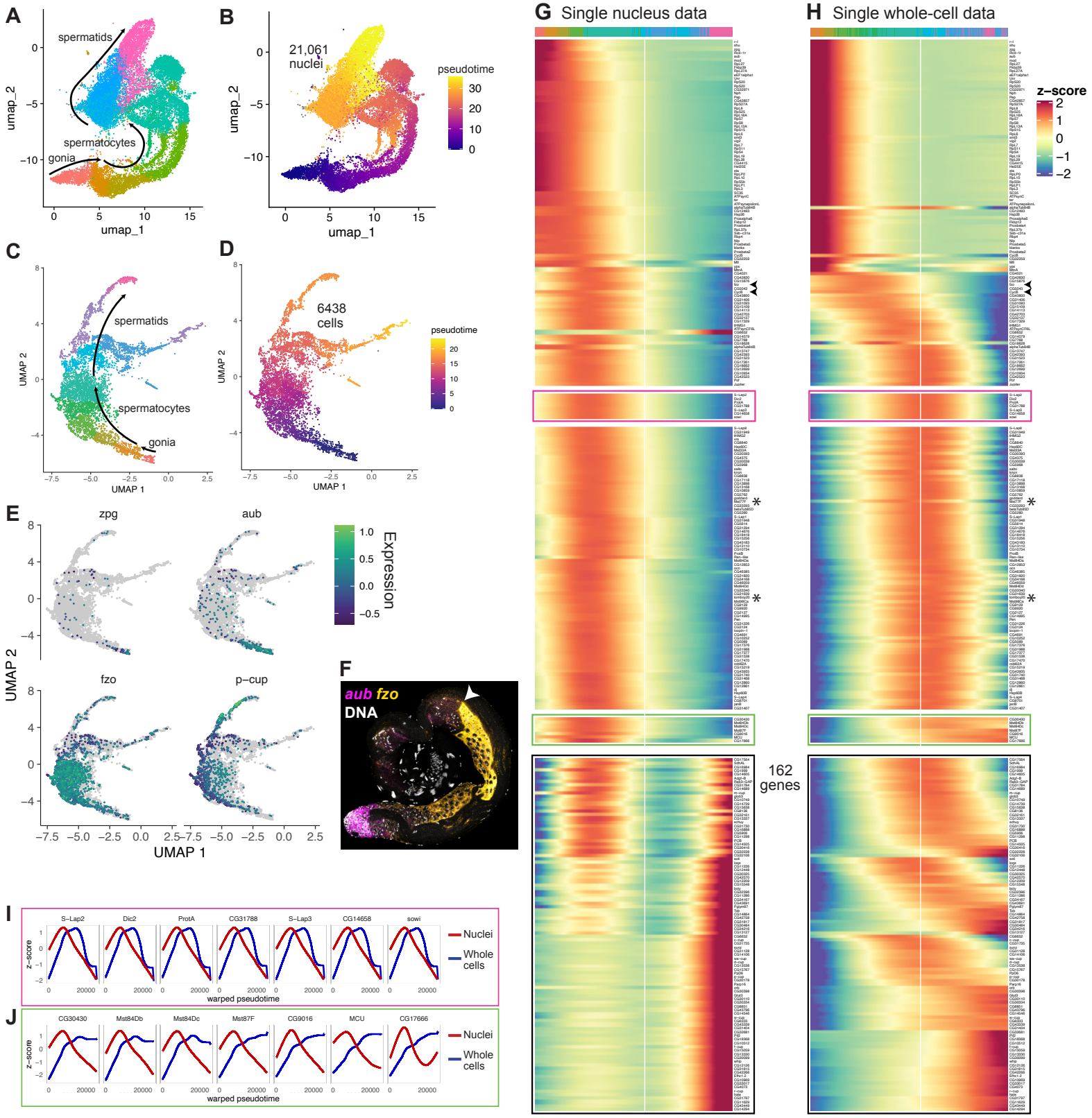
A
spermatids
spermatocytes
gonia
umap_1
umap_2
B
21,061
nuclei
pseudotime
C
spermatids
spermatocytes
gonia
UMAP 1
UMAP 2
D
6438
cells
pseudotime
E
zpg
aub
fzo
p−cup
Expression
F
aub fzo
DNA
G Single nucleus data
H Single whole-cell data
z−score
162
genes
I
S−Lap2
Dic2
ProtA
CG31788
S−Lap3
CG14658
sowi
Nuclei
Whole
cells
z−score
warped pseudotime
J
CG30430
Mst84Db
Mst84Dc
Mst87F
CG9016
MCU
CG17666
Nuclei
Whole
cells
z−score
warped pseudotime

Figure 4

Developmental transitions revealed by comparing sn and scRNA-seq

(A-D) UMAP plots of germline-annotated data from **A,B)** FCA snRNA-seq of adult testis plus seminal vesicle and **C,D)** whole-cell scRNA-seq of adult testis. **A,C)** Color denotes germline differentiation stage. **B,D)** Color denotes pseudotime, with the few nuclei lacking a calculable pseudotime value colored gray. **E)** UMAPs of scRNA-seq data showing log10(Expression) levels of cell-stage diagnostic markers *zpg*, *aub*, *fzo*, and *p-cup*. **F)** FISH of diagnostic genes *aub* and *fzo*. Arrowhead marks end of *fzo* expression in early round spermatids. **(G-H)** Heatmaps of row-normalized (z-score) gene expression over pseudotime for: **G)** all germline-annotated single nuclei from panel A, **H)** all germline-annotated single whole cells from panel C, with genes in same order as in G. X axes, pseudotime; Y axes, genes. Vertical white line: nuclei (G) or cells (H) where level of fzo mRNA has dropped half way (0 on Z score). Top bars: cell identity for each column, colored as in panels A,C. Black boxes: genes transcribed post-meiotically (see Figure 5). **(I-J)** Comparison of gene expression over warped pseudotime for: **I)** genes outlined by pink boxes in G,H; **J)** genes outlined by green boxes in G,H.

350 Trajectory inference can assign a differentiation distance parameter to cells inferred from
351 transcriptional differences, with distance noted as 'pseudotime' (Trapnell et al., 2014). Applying
352 trajectory inference independently to the snRNA-seq and scRNA-seq germ line datasets
353 produced contiguous trajectories. Using Monocle3, 99.9% (21,061/21,091) of the snRNA-seq
354 germline nuclei were connected (Figure 4B). Notably, unlike prior trajectory analysis using
355 Slingshot (Li et al., 2022), the inferred trajectory was contiguous, connecting cells of all
356 differentiation points from early spermatogonia to late spermatids. Likewise, Monocle3 analysis
357 of the 6438 germline cells from the scRNA-seq also produced a contiguous trajectory from
358 spermatogonia to elongating spermatids, although it did include a late bifurcation, the explanation
359 of which may be technical or biological. For both datasets, pseudotime staging paralleled the
360 ordered trajectory deduced from marker gene expression in UMAP clusters. (Figure 2, Figure 4A-
361 E).

362 Plotting normalized gene expression across pseudotime in the sn- and scRNA-seq
363 datasets revealed both shared and contrasting dynamics. In both datasets, the same set of genes
364 were expressed in early germ cells, with expression diminishing over pseudotime (Fig 4G,H;
365 red/orange in the top color bar, corresponding to colors in Figure 4A,C). Similarly, some genes,
366 including *fzo* and *CycB* (arrowheads) reached their peak expression in spermatocytes and
367 dropped to low levels by early spermatid stages in the scRNA-seq (see vertical white line, Figure
368 4G,H), consistent with published *in situ* hybridization data (White-Cooper et al., 1998 and Figure
369 4F).

370 Interestingly, in the snRNA-seq dataset, a group of over 200 genes in the middle region
371 of the heat map reached peak expression in the mid-spermatocyte stages (green-aqua hues in
372 the top color bar) then dropped in expression, falling halfway to their nadir at a point similar to the
373 drop in *fzo* and *cycB* expression (vertical white line) (Figure 4G). In the scRNA-seq dataset,
374 however, these same genes were still at or near peak mRNA accumulation (red) in the same cells
375 in which expression of *fzo* and *cycB* had already dropped (Figure 4H). Thus it is the comparison
376 of the datasets that is most revealing: these genes are transcribed in spermatocytes, then
377 transcription halts (inferred from the snRNA-seq dataset), but their mRNAs remain high in

378 spermatids (inferred from the scRNA-seq dataset) well past the stage when *fzo* and *cycB* mRNAs
379 disappear (inferred from both datasets). Several genes from this set, including *Male-specific*
380 *transcript 77F* (*Mst77F*) and *tomboy20* (asterisks), have been previously demonstrated by *in situ*
381 hybridization to maintain abundant transcripts in both spermatocytes and elongating spermatids
382 (Barckmann et al., 2013; Hwa et al., 2004). Comparison of the snRNA-seq and scRNA-seq
383 datasets suggests that such a pattern, previously described for a small number of transcripts, is
384 shared by hundreds of genes.

385 Comparison of single cell and single nucleus data revealed several distinct classes of
386 transcript behaviors in spermatids, each worthy of targeted follow-up study. Notably, perduring
387 transcripts from the class of genes described above showed two types of behavior. For most
388 genes, transcripts disappeared sharply in later elongating spermatids (examples in pink box in
389 Figure 4G,H). To plot transcript levels in the sn- and scRNA-seq datasets on the same X axes,
390 a common 'warped' time scale was derived for the datasets (see Materials and Methods). First,
391 for all genes upregulated in spermatocytes, onset of transcription in the nuclear transcriptome
392 was followed (with delay) by upregulation in the whole-cell transcriptome, generally dominated by
393 cytoplasmic transcripts (Figure 4I,J). As seen in the heat map (Figure 4H), the graphs for the pink
394 box genes show that the mRNAs remained at peak levels considerably later in the scRNA-seq
395 than in the snRNA-seq data, but the transcripts were eventually strongly downregulated by late
396 spermatid stages (Figure 4I: blue lines). This suggests complexity in mRNA regulation in the
397 cytoplasm: stable maintenance in early spermatids and abrupt degradation in later spermatids,
398 perhaps once transcripts have been translated. Interestingly, the protein products of several of
399 these genes are present and functional in late spermatids and sperm (Raja and Renkawitz-Pohl,
400 2006), suggesting these proteins are actively maintained in the absence of new translation.

401 A second type of behavior was noted for a small group of genes (green box in Figure
402 4G,H, graphed in J) where transcripts perdured even longer, remaining high through the latest
403 stages assessed by scRNA-seq (blue lines in Figure 4J). As differentiation of late spermatocytes
404 to late spermatids takes days (Chandley and Bateman, 1962), these remarkable transcripts
405 maintained high levels of cytoplasmic abundance, with almost no sign of degradation, even days

406 after active transcription had dropped off. This suggests exceptional stability, likely provided by
407 specialized RNA-binding proteins. Some such transcripts, encoded by *Mst84Db*, *Mst84Dc*, and
408 *Mst87F*, have long been recognized to be translationally regulated, with perdurance in the
409 cytoplasm for up to three days (Kuhn et al., 1991; White-Cooper et al., 1998). Others, including
410 *CG30430, CG9016, MCU*, and *CG17666* have not been previously reported to undergo
411 translational regulation. The differences in degradation timing revealed by scRNA-seq (Figure
412 4G-J, pink vs green boxes) may hint that distinct groups of RNAs, and thus their protein products,
413 are engaged at different stages of spermatid morphogenesis.

414 Another compelling example of the utility of comparing snRNA-seq and scRNA-seq data
415 is highlighted by the group of genes outlined in black (Figure 4G,H). In the snRNA-seq dataset,
416 these genes are expressed in spermatocytes but transcription shuts down in early spermatids
417 and remains off for a considerable period before expression is activated again in mid-to late
418 elongation stages. Thus, few transcripts bridge the gap between late spermatocytes and mid-
419 stage elongating spermatids, as if the two stages were disconnected. In contrast, in the scRNA-
420 seq dataset, many of these same genes showed continued high transcript levels throughout the
421 spermatid stages, presumably representing storage of mRNAs in the cytoplasm (Figure 4G,H,
422 black boxes). In consequence, the mature spermatocyte to elongating spermatid stage
423 transcriptomes were well connected through a smooth gradient of transcript levels in scRNA-seq
424 data.

425 **Reactivation of transcription in mid-to-late elongating spermatids**

426 The ability of snRNA-seq to highlight dynamic transcriptional changes during cellular
427 differentiation revealed striking transcriptional (re)activation of a subset of 162 genes in mid-to-
428 late elongating spermatids, a phenomenon previously described for only 24 genes, "called post-
429 meiotic transcripts" (Barreau et al., 2008). UMI counts and average number of genes expressed
430 in the snRNA-seq were relatively low in post-meiotic clusters compared to other germ cell clusters
431 (Figure 3C), consistent with the long-held idea that spermatids are nearly transcriptionally silent.
432 Nonetheless, post-meiotic transcription appeared more extensive than previously appreciated,
433 with transcripts from approximately 1000 genes detected. Analysis for genes specifically

Figure 5

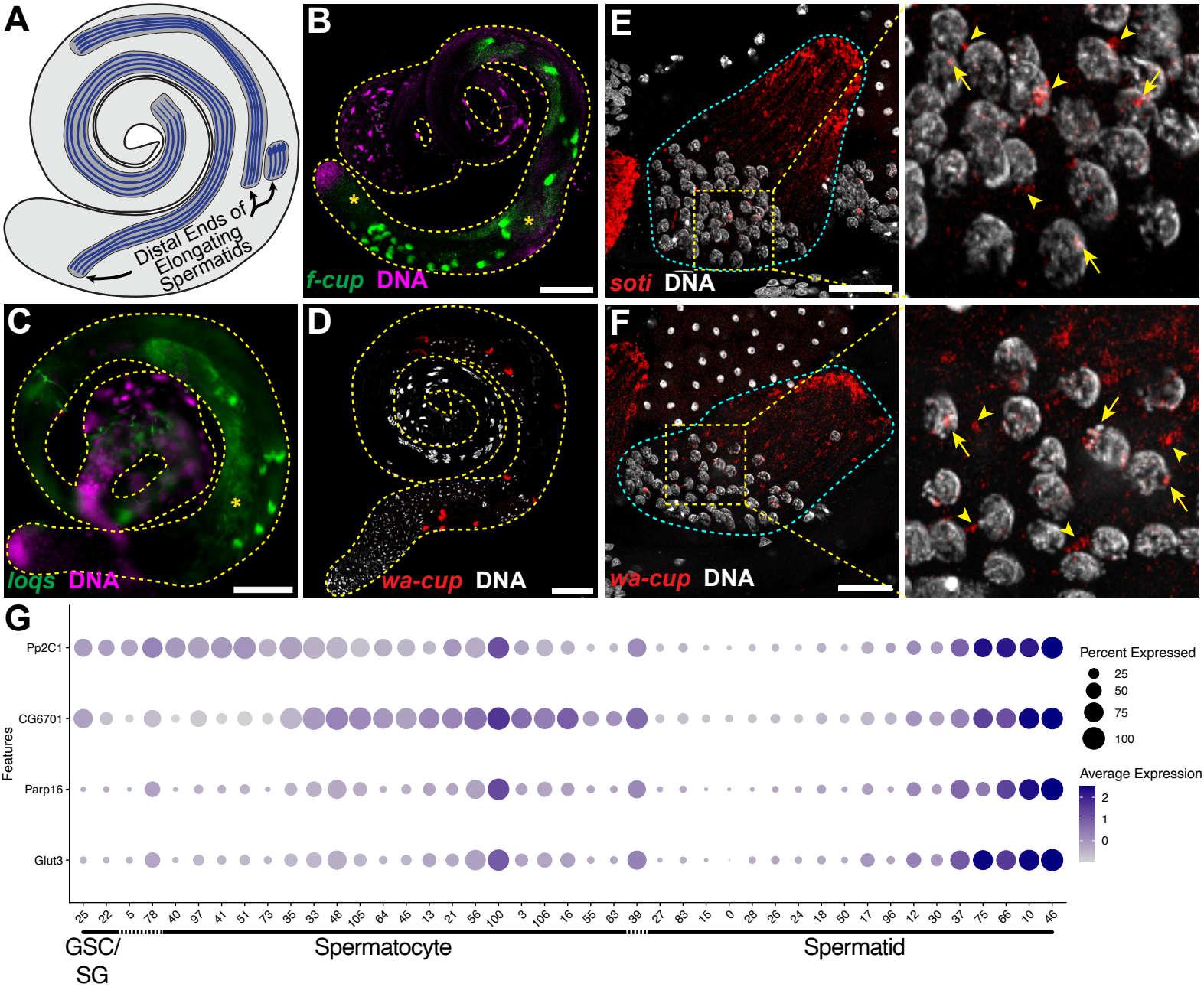
A
Distal Ends of Elongating Spermatids
B
f-cup DNA
E
soti DNA
C
loqs DNA
D
wa-cup DNA
F
wa-cup DNA
G
Pp2C1
CG6701
Parp16
Glut3
Features
Percent Expressed
25
50
75
100
Average Expression
2
1
0
25 22 5 78 40 97 41 51 73 35 33 48 105 64 45 13 21 56 100 3 106 16 55 63 39 27 83 15 0 28 26 24 18 50 17 96 12 30 37 75 66 10 46
GSC/ SG
Spermatocyte
Spermatid
H
Pp2C1
CG6701
Parp16
Glut3
Expression
Fold Change
4.03
1.74
8.65
7.65
pseudotime

Figure 5 with one supplement

The transcript landscape of haploid spermatids

A) Diagram of spermatid orientation in the testis at different stages of spermatid elongation. Arrows: distal ends of spermatid cysts. (**B,C)** RNA FISH of representative transcripts (**B**: *f-cup*, **C**: *loqs*) in whole testes showing different patterns of mRNA localization for post-meiotically transcribed genes. DNA (magenta), target RNA (green). Asterisk: signal in spermatocytes. Bar: 100µm. **D)** smFISH for *wa-cup* (red) in whole testes. DNA (white). Bar: 100µm. (**E,F)** Left: smFISH for *soti* **E**) or wa-cup **F**) in a single early elongating spermatid cyst (cyan dashed line). RNA (red) and DNA (white). Bar: 25µm. Right: Enlarged image of yellow dashed box showing spermatid nuclei. Arrows: nuclear transcripts. Arrowheads: perinuclear granules. **G)** Dot plot for selected spermatid transcribed genes showing expression levels in each germ cell cluster. **H**) Expression over pseudotime for selected spermatid transcribed genes and fold change between late pseudotime and previous expression maxima in early/mid pseudotime. (see also FIgure 5 - figure supplement 1).

434 enriched in late pseudotime identified a list of 162, here termed spermatid transcribed genes
435 (Figure 4G, black box, and Figure 5 - source data 1). These included 18 of the previously
436 identified 24. FISH revealed *flyers-cup* (*f-cup*) RNA at the distal end of elongated spermatid
437 bundles, as expected (Barreau et al., 2008, Figure 5A,B; Figure 5 - figure supplement 1A). RNA
438 from *loquacious (loqs)*, a newly identified spermatid transcribed gene, was similarly localized
439 (Figure 5C; Figure 5 - figure supplement 1B). Transcripts from *walker cup* (*wa-cup*) and *scotti*
440 (*soti*) also localized to the distal ends of elongating spermatids as expected (Barreau et al., 2008,
441 Figure 5D-F; Figure 5 - figure supplement 1C, D). Analysis of earlier elongating spermatid cysts
442 by single molecule RNA FISH (smFISH) supports active transcription of *wa-cup* and *soti* in
443 spermatids: smFISH revealed foci in spermatid nuclei, suggesting nascent, post-meiotic
444 transcription (Figure 5E,F, arrows), as well as perinuclear granules (arrowheads), which could
445 represent newly synthesized RNAs being trafficked toward the distal ends of the spermatids.

446 Analysis of the snRNA-seq data showed that many newly identified spermatid transcribed
447 genes, including *Pp2C1* and *CG6701*, were initially expressed in spermatocytes or
448 spermatogonia, downregulated in early spermatids, and later reactivated during mid-to-late
449 elongation (Figure 4F-upper half of black box; Figure 5G,H). Other newly identified spermatid
450 transcribed genes, including *Parp16* and *Glut3,* were weakly expressed in spermatocytes but
451 robustly transcribed in elongating spermatids (Figure 4F - lower half of black box; Figure 5G,H).
452 Both patterns are consistent with RT-qPCR and RNA *in situ* hybridisation results for the 24 post-
453 meiotic transcripts previously identified (Barreau et al., 2008). Together the results show two
454 sources of RNAs in elongating spermatids: cytoplasmic perdurance of RNAs transcribed in
455 spermatocytes (Figure 4H), and *de novo* post-meiotic (re)activation of transcription of certain
456 genes (Figure 4G, Figure 5G,H).

457 The majority of the spermatid transcribed genes remain functionally uncharacterized, and
458 await investigation. GO term analysis showed no significant enrichment for any single biological
459 process or pathway, although several functional classes were represented (Figure 5 - source data
460 1). Additionally, genes in this set did not appear to be coordinately reactivated, as by a single
461 regulatory circuit. Rather, the likelihood of genes to be (re)activated concordantly was weakly
462 correlated with their expression level, with a few outliers (Figure 5 - figure supplement 1E).

463

464 **Progression of differentiation in the somatic cyst cell lineage**

465 Somatic cyst cells govern many germline transitions, from stem cell behavior through
466 sperm maturation and release (Figure 1A, 6O). The snRNA-seq approach may be especially
467 useful for characterizing cyst cell transcriptomes across differentiation stages because the long,
468 thin, extended shape of many cyst cell types may make isolation of intact cells difficult.

469 Cyst lineage identity was assigned by expression of three transcriptional regulators, *traffic*
470 *jam* (*tj*), *eya* and *sine oculis* (*so*) (Figure 1B; Figure 6A,B,V; Figure 6 - figure supplement 1C,D).
471 While both very early clusters (62, 36, 58) and very late clusters (72, 79, 84) were simple to
472 describe (see below), the middle region of the 2-D UMAP presented a tangle. However, re-
473 projecting the lineage to preserve a third dimension clarified the assignment of cluster order
474 through the middle region (Figure 6C; Materials and Methods for a rotatable 3-D version). This,
475 combined with pseudotemporal ordering (Figure 6 - figure supplement 1A,B) and marker analysis,
476 enabled us to assign specific cyst cell clusters to stem cell, spermatogonial, spermatocyte,
477 spermatid or sperm release stages of germline development. The few clusters difficult to assign,
478 perhaps notably, had relatively low UMI count (FIgure 6E).

479 The cyst lineage begins with proliferative *stg*-expressing CySCs (Cluster 62, Figure 1D;
480 Figure 6A,D). Expression of *tj* suggested that clusters 36 and 58 represent early-stage cyst cells
481 that enclose transit amplifying spermatogonia (Figure 6B,V) labeled SgCC in Figure 6D, as Tj
482 protein marks the nucleus of these early cyst cells but is not detected by immunofluorescence
483 staining in spermatocyte-associated cyst cells (Li et al., 2003; Zoller and Schulz, 2012).
484 Complementing this, *CG3902* was also detected in the same nuclei as *tj*, and a CG3902 protein
485 trap line revealed cytoplasmic protein accumulation up to early ScCC and no detectable protein
486 thereafter (Figure 6V; Figure 6 - figure supplement 1E,F). *P-element induced wimpy testis* (*piwi*)
487 mRNA was enriched in the same clusters as *tj and CG3902*, and also was detected at lower levels
488 in nuclei of subsequent clusters (for example, clusters 47 and 77; Figure 6D SgCC and ScCC,
489 F,V) (Gonzalez et al., 2015). FISH and analysis of a Piwi protein trap confirmed expression in
490 early cyst cells associated with spermatogonia at the testis apex, as well as in cyst cells enclosing

Figure 6

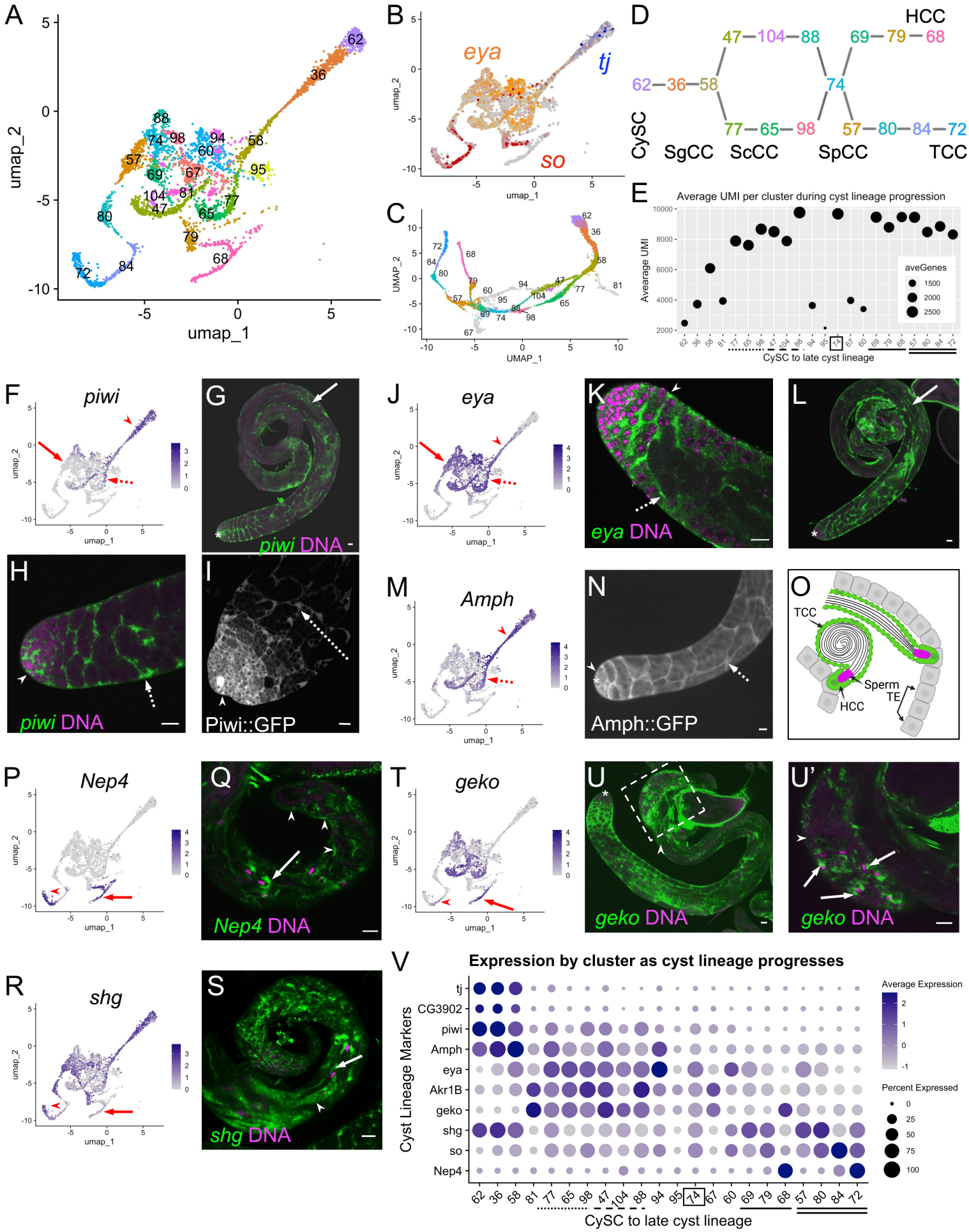
A
umap_2
umap_1
B
eya
tj
so
umap_2
umap_1
C
UMAP_2
UMAP_1
D
HCC
CySC
SgCC
ScCC
SpCC
TCC
E
Average UMI per cluster during cyst lineage progression
Avearage UMI
aveGenes
CySC to late cyst lineage
F
piwi
G
piwi DNA
H
piwi DNA
I
Piwi::GFP
J
eya
K
eya DNA
L
M
Amph
N
Amph::GFP
O
TCC
Sperm
TE
HCC
P
Nep4
Q
Nep4 DNA
T
geko
U
geko DNA
U'
geko DNA
R
shg
S
shg DNA
V
Expression by cluster as cyst lineage progresses
Cyst Lineage Markers
tj
CG3902
piwi
Amph
eya
Akr1B
geko
shg
so
Nep4
Average Expression
Percent Expressed
CySC to late cyst lineage

Figure 6 with one supplement

Differentiation in the somatic cyst cell lineage

A) Cyst cell lineage portion of the UMAP from snRNA-seq data with Leiden 6.0 cluster numbering. **B)** Expression of *tj* (blue), *eya* (orange), and *so* (red) projected on the UMAP (heatmaps in FIgure 6 - figure supplement 1C,D). **C)** 2D UMAP of cyst cell clusters newly reprojected in their own gene expression space (Materials & Methods; note different axis coordinates relative to panel A). Cluster colors correspond to panel A except for unidentifiable clusters with low UMIs (gray). **D)** Schematic of cyst cell cluster progression from CySC, to spermatogonia-associated (SgCC) to spermatocyte-associated (ScCC) to Spermatid-associated (SpCC) inferred from the 3D UMAP in panel C. Cluster numbers and colors as in A. Note two splits, one earlier and one later in cyst lineage progression. **E)** Plot of average number of genes (dot size) and average number of UMIs per cyst lineage-annotated cluster; with clusters ordered by progression of differentiation. Dotted and dashed lines under cluster numbers represent the early split shown in D, while single (HCC) and double (TCC) solid lines represent the later split. **F - U)** In these panels, a UMAP is compared to a set of FISH and/ or IF images. For each gene comparison, arrows, dashed arrows and/or arrowheads point out the same cell type in the UMAP and its corresponding FISH and/ or IF image. **F)** *piwi* mRNA expression projected on the UMAP. **(G,H)** *piwi* mRNA (FISH; green) and DNA (magenta). **G**) whole testis, and **H)** testis tip view. **I)** Apical tip of testis expressing GFP protein trap of Piwi. **J)** *eya* mRNA expression projected on the UMAP. **(K, L)** *eya* mRNA (FISH; green) and DNA (magenta), **K**) Testis apical tip. **L)** Whole testis. **M)** *Amph* expression projected on the UMAP. **N)** Testis apical third expressing GFP protein trap of Amph**.** **O)** Schematic of the head cyst cell (HCC, green solid outline) embedded in the terminal epithelium (TE, gray), with the tail cyst cell (TCC, green dashed outline) extending away, containing either a pre-coiled (top) or coiled (bottom) spermatid bundle (nuclei, magenta). **P)** *Nep4* mRNA expression projected on the UMAP. **Q)** Testis base showing *Nep4* mRNA (FISH, green) and DNA (magenta). **R)** *shg* mRNA expression projected on the UMAP. **S)** Testis base showing *shg* mRNA (FISH, green) and DNA (magenta). **T)** UMAP of *geko* mRNA expression. **U)** Whole testis showing *geko* mRNA (FISH, green) and DNA (magenta). **U')** Inset outlined by white dashed box in U showing base of testis. **V)** Dotplot of gene expression (Y axis) by cluster as cyst lineage progresses through differentiation (X axis, left to right). Degree of expression in each cluster indicated by color scale. Percent of

cells within a cluster expressing the gene indicated by size of dot. Lines under cluster numbers as in E. Asterisk denotes hub. Bars: 20μm. (see also Figure 6 - figure supplement 1).

491 growing spermatocyte cysts (Figure 6G-I, arrowhead and dashed arrow, respectively).
492 Interestingly, *piwi* transcripts were also detected in more mature cyst cells associated with
493 elongated and polarized spermatids (Figure 6G, solid arrow), highlighting the differences between
494 active transcription detected by snRNA-seq and perdurance of cytoplasmic RNA detected by
495 FISH.

496 Both *tj* and *eya* mRNA expression was detected in cluster 58, but onward *tj* mRNA was
497 abruptly down regulated while *eya* transcript expression increased. We surmise this marks the
498 transition to cyst cells associated with spermatocytes (Figure 6D,J,V), since Eya protein is known
499 to accumulate in cyst cell nuclei from late stage spermatogonia to spermatocytes (Fabrizio et al.,
500 2003). FISH confirmed the *eya* pattern (Figure 6K, arrowhead and dashed arrow), while also
501 revealing accumulation of *eya* transcript in cyst cells associated with post-meiotic spermatids
502 (Figure 6L, solid arrow).

503 Intriguingly, the cyst lineage bifurcates after cluster 58, with clusters 77, 65, and 98
504 successively in one arm and 47, 104, and 88 in the other (Figure 6C,D). This might be due to the
505 onset of differentiation of head versus tail cyst cells, and would represent the first hint at when
506 this occurs in the lineage. Identification and characterization of genes differentially expressed within
507 the split could reveal whether cyst cells specific to a given arm of the lineage govern different
508 properties of spermatocyte-containing cysts.

509 Expression of *Amphiphysin* (*Amph)* supported the conclusion that the bifurcation in the cyst
510 cell lineage after cluster 58 represents cyst cells associated with spermatocytes. *Amph* mRNA
511 was high before the bifurcation and persisted at lower levels in the two arms of the split, dropping
512 substantially by cluster 74 (Figure 6M,V). Expression of a protein trap confirmed that *Amph* protein
513 levels were high in SgCCs and ScCCs (Figure 6N, arrowhead and dashed arrow, respectively)
514 and declined significantly in cysts containing early spermatids (data not shown). It is intriguing
515 that the snRNA-seq and protein trap indicate that Amph expression is strongest in early cyst cells,
516 even though it encodes a BAR domain protein required to form the actomyosin clamp that
517 maintains head cyst cell membrane integrity as these cells wrap around spermatid heads late in
518 spermatogenesis (Kapoor et al., 2021).

519 Branches of the cyst lineage rejoin at cluster 74 implying a transition to a common
520 transcriptional state. Interestingly, comparison of *Akr1B* vs. *Amph* suggests that cluster 74
521 contains cyst cells associated with early or elongating spermatids. Akr1B transcripts were
522 elevated in both arms of the split through to cluster 74 (Figure 6 - figure supplement 1G) while
523 *Amph* expression had decreased by cluster 74 (Figure 6M,V). Analysis of a *Akr1B::GFP* protein
524 trap confirmed its expression in cyst cells associated with elongating spermatid cysts (SpCC)
525 (Figure 6 - figure supplement 1H,I, arrows). Thus, it appears that, with respect to the
526 transcriptome, a developmental transition occurs within the cyst lineage as these cells mature
527 from support of spermatocytes to early spermatid cysts.

528 After cluster 74, the lineage again splits, with marker analysis suggesting that this correlates
529 with differentiation of late stage head cyst cells (HCC, clusters 69, 79, 68) versus tail cyst cells
530 (TCC, clusters 57, 80, 84, 72; Figure 6A,C,D). None of the clusters along either late branch were
531 enriched for *eya* (Figure 6J,V). However, *Neprilysin-4* (*Nep4*), a metalloprotease involved in male
532 fertility (Sitnik et al., 2014), was upregulated in both late branches (Figure 6P,V), and FISH
533 confirmed *Nep4* expression in late HCCs and TCCs associated with fully coiled spermatids
534 (Figure 6Q, arrow and arrowheads, respectively). Intriguingly, the snRNA-seq data showed
535 expression of the E-cadherin *shotgun* (*shg*) throughout the cyst lineage, with the notable
536 exception of cluster 68 (Figure 6A,R,V). Consistent with this cluster representing late HCCs, FISH
537 revealed higher levels of *shg* in late TCCs than in late HCCs (Figure 6S, arrowhead and arrow,
538 respectively). Conversely, the snRNA-seq data showed that *geko*, an olfactory gene not studied
539 in the testis (Shiraiwa et al., 2000), was upregulated both in cyst cells associated with
540 spermatocytes and in part of cluster 68 (Figure 6A,T,V). In fact, higher resolution (Leiden 8.0)
541 analysis divided cluster 68 into nuclei either enriched or not for *geko* (data not shown). Although
542 FISH for *geko* revealed expression in cyst cells associated with spermatocytes and elongating
543 spermatid cysts throughout the testis (Figure 6U), close inspection of the testis basal region
544 showed high expression in HCCs (Figure 6U', arrows) and lower expression in TCCs (Figure
545 6U,U', arrowhead) as predicted by the UMAP (Figure 6T). Exploring the HCC and TCC clusters
546 further will be revealing given the unique roles played by these cells during spermatid retraction
547 and coiling (Figure 6O).

548 ## The hub: architectural organizer and key signaling center

549 The hub is a small group of somatic, epithelial-like cells at the testis apex that acts as a
550 niche, providing signals that maintain GSC and CySC fate (Hardy et al., 1979; Kawase et al.,
551 2004; Kiger et al., 2001; Leatherman and DiNardo, 2010; Shivdasani and Ingham, 2003; Tulina
552 and Matunis, 2001). Initial marker analysis suggested that the hub maps to cluster 90. However,
553 only 79 of these 120 nuclei clustered together, while other members of cluster 90 were dispersed
554 (Figure 7 - figure supplement 1A). Additionally, the outcast nuclei either expressed germline
555 genes such as *Rbp4, zpg, p53*, and *vas* (Figure 7 - figure supplement 1B-E) or did not consistently
556 express signature genes known to be enriched in hub cells (Figure 7A-D, dashed red circles),
557 strongly suggesting they differ transcriptionally from the 79 tightly clustered nuclei. Consequently,
558 we pared cluster 90 down to 79 definitive hub nuclei (see Materials and Methods).

559 The snRNA-seq data show that hub nuclei express genes involved in signaling as well as
560 markers common in epithelial cells. For example, *upd1* and *hedgehog* (*hh)* are upregulated,
561 consistent with the hub's role in stem cell maintenance (Figure 7A,B,K) (Amoyel et al., 2013; Kiger
562 et al., 2001; Michel et al., 2012; Tulina and Matunis, 2001). Additionally, hub cells are enriched
563 for proteins implicated in junctional complexes, including factors such as Fasciclin III (Fas3), E-
564 Cadherin (Shg), N-Cadherin (CadN), Discs large (Dlg1), and Cheerio (Cher; an orthologue of
565 Filamin) (Boyle et al., 2007; Le Bras and Van Doren, 2006; Brower et al., 1981; Papagiannouli
566 and Mechler, 2009; Tanentzapf et al., 2007). In agreement with this, hub nuclei exhibited
567 significant expression of these markers (Figure 7C,D,K).

568 The definitive identification of hub nuclei allowed analysis for upregulated genes ($\log_2$FC
569 ≥ 1, compared to the full testis plus seminal vesicle dataset; Figure 7 - source data 1) as new
570 candidate hub markers. One such gene encodes the transcription factor Tailup (Tup, also known
571 as Islet (Boukhatmi et al., 2012; She et al., 2021; Thor and Thomas, 1997) (Figure 7E). Indeed,
572 a Tup::GFP transgene showed strong protein expression marking hub nuclei at the apex of adult
573 testes (Figure 7F-F'''). This is consistent with recent evidence that Tup is expressed in and
574 required for niche organization in the male embryonic gonad (Anllo and DiNardo, 2022; Anllo et
575 al., 2019).

Figure 7

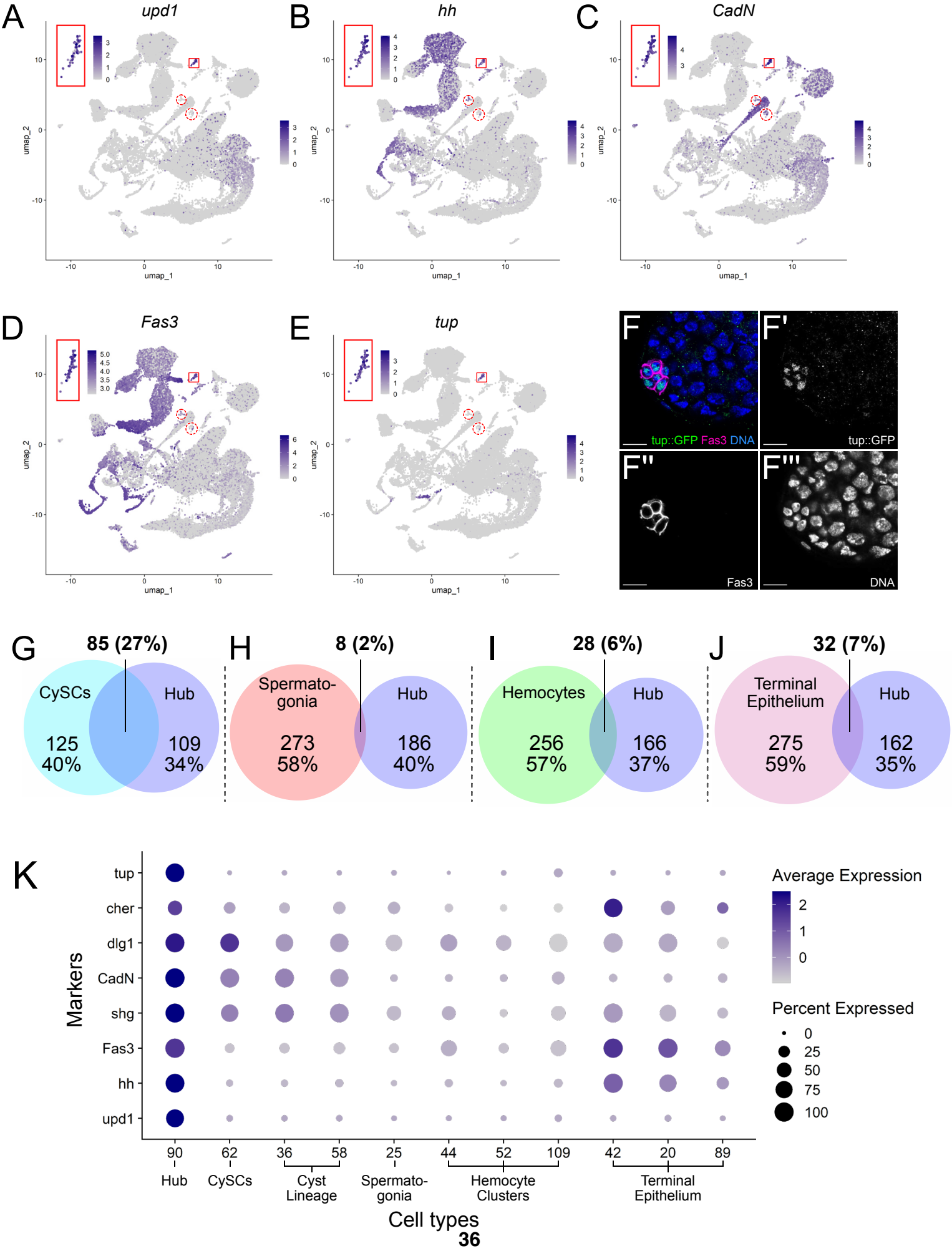
A
upd1
B
hh
C
CadN
D
Fas3
E
tup
umap_1
umap_2
F
tup::GFP Fas3 DNA
F'
tup::GFP
F''
Fas3
F'''
DNA
G
85 (27%)
CySCs
Hub
125
40%
109
34%
H
8 (2%)
Spermato-
gonia
Hub
273
58%
186
40%
I
28 (6%)
Hemocytes
Hub
256
57%
166
37%
J
32 (7%)
Terminal
Epithelium
Hub
275
59%
162
35%
K
Markers
tup
cher
dlg1
CadN
shg
Fas3
hh
upd1
Average Expression
Percent Expressed
90
62
36
58
25
44
52
109
42
20
89
Hub
CySCs
Cyst
Lineage
Spermato-
gonia
Hemocyte
Clusters
Terminal
Epithelium
Cell types

Figure 7 with one supplement

Characteristics of the Hub

(A-E) Expression of the indicated genes (*upd1*, *hh*, *CadN*, *Fas3*, *tup*) projected onto the testis snRNA-seq UMAP with the 79 definitive hub nuclei outlined (small red box). Color intensity corresponds to expression level, shown as normalized average logFC. A reprojection of hub nuclei is shown for each panel (larger red box) with its own key for expression. Red dashed circles contain non-hub nuclei of cluster 90 (see text and Figure S5). **F-F'''**) Apical tip of adult testis carrying a Tup::GFP fusion transgene revealing expression largely restricted to hub nuclei delimited by Fas3 (magenta). Nuclei (blue). Scale bar is 10μm. **G-J)** Paired Venn diagrams, comparing up-regulated genes in the hub vs. clusters containing either CySCs, spermatogonia, hemocytes or terminal epithelial cells. Overlap in upregulated genes greatest between Hub and CySCs. In each pairing, circle size reflects the number of genes compared for each cluster. Genes are listed in Figure 7 - source data 1; the specific genes for hub vs. CySCs in Figure 7 - source data 2. **K)** Dot plot of expression of selected genes comparing hub to CySCs, the early Cyst lineage, Spermatogonia, Hemocytes, and Terminal epithelium. (see also Figure 7 - figure supplement 1).

576 Previous lineage-tracing showed that hub cells and CySCs derive from a common pool of
577 gonadal cells during embryogenesis (Le Bras and Van Doren, 2006; DeFalco et al., 2008;
578 DiNardo et al., 2011). Comparison of genes up-regulated in hub cells with those up-regulated in
579 CySCs strongly reflects this developmental relatedness. Hub and CySC (cluster 62) nuclei share
580 27% of their up-regulated genes, likely reflecting their embryonic co-origin (Figure 7G, Figure 7 -
581 source data 2). In contrast, the fraction of shared upregulated genes was much lower between
582 the hub and several lineally and functionally distinct cell types, including spermatogonia,
583 hemocytes and terminal epithelia (Figure 7 H-J). Supporting this, several genes up-regulated in
584 the hub are also highly expressed in the CySCs and the cyst lineage, but much lower in
585 spermatogonia and hemocytes (Figure 7K). The transcriptional similarity observed in the adult
586 cell types could account for the ability of hub cells to replenish CySCs after drastic injury, as well
587 as explain the shift of one lineage toward the other when specific gene functions are compromised
588 (Greenspan et al., 2022; Herrera et al., 2021; Hétié et al., 2014; Okegbe and DiNardo, 2011; Voog
589 et al., 2014).

590 **Epithelial Cells of the Testis Organ**

591 A key role of the terminal epithelial cells (TE) is to anchor the head end of elongated
592 spermatid bundles at the base of the testis during individualization and coiling (Figure 6O) so that
593 the sperm are positioned for release into the seminal vesicle (SV). Marker gene analysis
594 suggested that both the TE and SV reside in the "mermaid" of the UMAP (Figure 1B, 8A). snRNA-
595 seq showed enrichment for *hh* in two broad areas of the UMAP (Figure 8B; Figure 8 - figure
596 supplement 1A), while FISH showed *hh* RNA in TE and SV cells (Figure 8C, arrowhead and
597 arrow, respectively). *Metallothionein A* (*MtnA*) was most highly upregulated in the lower clusters
598 enriched for *hh* (Figure 8A,B,D; Figure 8 - figure supplement 1B; clusters 42, 20), with colorimetric
599 ISH confirming strong expression of *MntA* in the TE (Figure 8E, arrow). The SV marker *Neprilysin-*
600 *5* (*Nep5;* Sitnik et al., 2014), confirmed by FISH to be expressed in SV, was upregulated in the
601 uppermost clusters (Figure 8F,G, arrow; Figure 6 - figure supplement 1C; clusters 1, 4, 7, 9, 29,
602 32, 34, 93). Although they encode nearly identical proteins, *Nep4* and *Nep5* are expressed in
603 different cell types of the adult testis (HCC/TCC and SV, respectively; Figure 8H). Note that while
604 the TE and SV each map as a single-block cluster at lower clustering resolution (Figure 2 figure
605 supplement 1B), the appearance of multiple clusters for each cell type at higher resolution

Figure 8

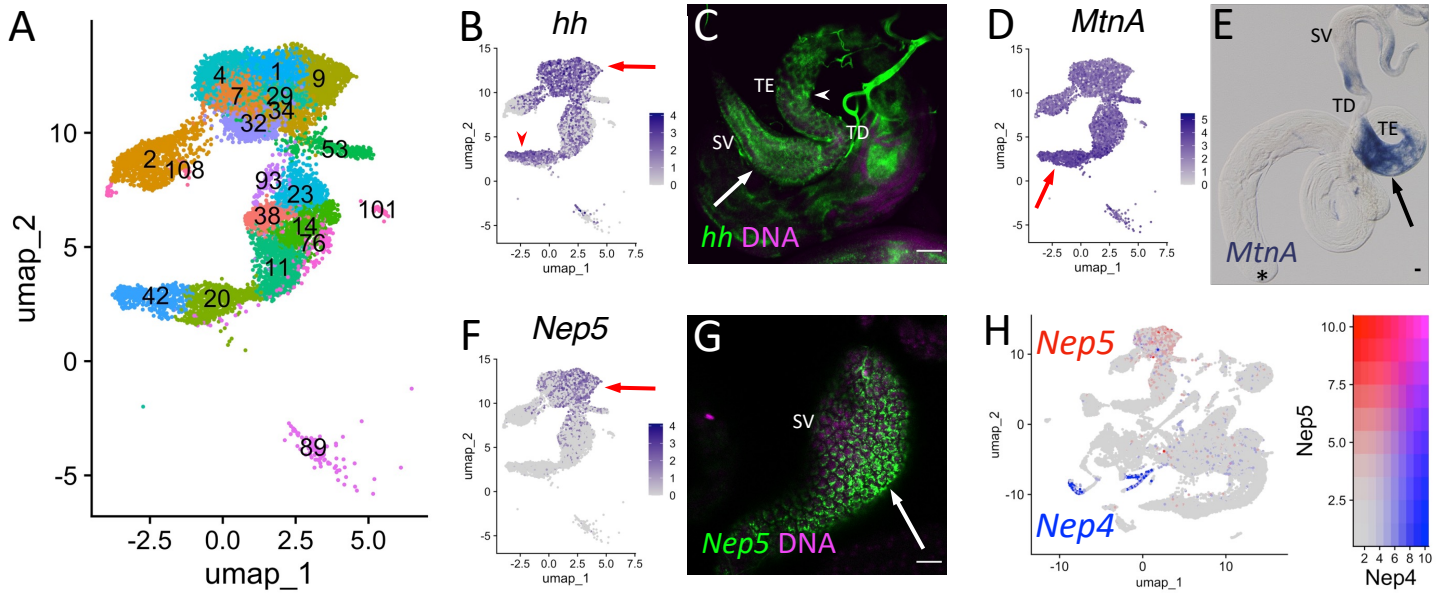
A
umap_2
umap_1
B
hh
C
TE
SV
TD
hh DNA
D
MtnA
E
SV
TD
TE
MtnA
F
Nep5
G
SV
Nep5 DNA
H
Nep5
Nep4
Nep5
Nep4

Figure 8 with one supplement

Characteristics of supporting epithelia

A) UMAP of non cyst cell lineage epithelial cells of the testis from the FCA snRNA-seq data with Leiden 6.0 clusters. Note: identity of cluster 89 remains undefined. **B - G)** In these panels, a UMAP is compared to a set of (F)ISH images. For each gene comparison, arrows and/or arrowheads point out the same cell type in the UMAP and its corresponding FISH image. **B)** *hh* mRNA expression projected on the UMAP. **C)** FISH of *hh* mRNA (green) and DNA (magenta) showing the base of the testis including the TE (arrowhead), testicular duct (TD), and SV (arrow). **D)** *MtnA* mRNA expression projected on the UMAP. **E)** Colorimetric *in situ* hybridization of *MtnA* mRNA (blue) in an entire testis. **F)** *Nep5* mRNA expression projected on the UMAP. **G)** FISH of *Nep5* mRNA (green) and DNA (magenta) showing the SV. **H**) *Nep4* (blue) and *Nep5* (red) mRNA expression projected on the UMAP with corresponding heatmap. Arrows and arrowheads point to corresponding cells on UMAPs and stained tissues. Asterisk denotes hub. Bars: 20μm. (see also FIgure 8 - figure supplement 1).

606 suggests notable transcriptional heterogeneity within both cell types (Figure 8A). Thus, epithelial
607 cell types in general represent an underexplored area in testis biology.
608
609

610 # Discussion

611 Study of the *Drosophila* testis has had great impact on reproductive biology and on
612 understanding stem cell-niche interactions, developmentally regulated cell cycles and cell type
613 specific cellular morphogenesis. Against this foundation, examining single cell and single nuclear
614 RNA-seq has revealed several notable features. First, in gene expression space, each stem cell-
615 based lineage was geographically isolated from the other, reflecting their distinct lineage identity.
616 Second, the individual cells within each stem cell lineage are essentially arranged in strings, due
617 to progressive changes in gene expression during differentiation. Third, a UMAP for each lineage
618 revealed complexities along its trajectory that reflect both known intricacies of development within
619 each lineage, as well as previously unappreciated gene expression dynamics. Fourth, in contrast
620 to the two stem cell lineages, terminally differentiated cells generally clustered into discrete
621 groups, as expected. Finally, the comparison of single cell and single nucleus datasets proved
622 particularly revealing of dynamic developmental transitions in lineage differentiation.
623

624 ## Germline and soma map to distinct gene expression domains

625 While *Drosophila* spermatogenesis relies on two separate stem cell lineages, the progeny
626 of each lineage associate intimately with each other, cooperating at multiple points during
627 differentiation to produce functional sperm. Despite their physical association, in gene expression
628 space the two lineages lie well-separated. This is not a surprise, since the lineages are specified
629 independently during embryogenesis in space and time, as well as by different gene regulatory
630 circuits. Such separation between germline and supporting soma is also observed in the scRNA-
631 seq from the *Drosophila* ovary and murine testis (Green et al., 2018; Rust et al., 2020).

632 Accurate pairing of each germline cluster with a somatic cluster representing its interacting
633 partner could facilitate identification of the underlying cell signaling circuits that form the basis for
634 cooperation between the lineages at various points along their differentiation trajectory. We are
635 able to highlight some likely pairing assignments due to the extensive knowledge of *Drosophila*

636 spermatogenesis and testis biology. For example, the CySC cluster (62) likely associates directly
637 with germ cells in cluster 25, while cyst cell lineage clusters 36 and 58 likely associate with later
638 spermatogonia (cluster 22). Other associations are also suggested by the data (Figure 6D) with
639 more remaining to be defined. The transcriptome in these pairings can be mined to identify
640 candidate signaling pathways by recently developed tools, such as FlyPhone (Liu et al., 2022).

641 **Stem cell lineages appear as strings along their differentiation path**

642 The data highlight how tissues maintained by stem cell lineages display a characteristic
643 geography in gene expression space. The UMAP reveals these lineages to be arranged largely
644 in strings, due to progressive changes in gene expression. The linear arrangement is strongly
645 supported by trajectory inference for both the sc- and sn-RNA-seq datasets (see also Li et al.,
646 2022). This arrangement might be diagnostic of at least some stem cell lineages, as it is observed
647 to a degree in germline data from murine testes, in murine small intestine (Green et al., 2018;
648 Haber et al., 2017), and in the follicle cell lineage in the *Drosophila* ovary (Li et al., 2022; Rust et
649 al., 2020). While in some tissues maintained by stem cells, such an arrangement is made
650 apparent only by trajectory inference, in the *Drosophila* testis it is apparent in the UMAP
651 representing gene expression space.

652 Additionally, even with progressive changes in gene expression along their respective
653 trajectories (Figures 2L, 6V), cells in these strings nevertheless can be sorted into clusters. The
654 extensive study of germline and somatic testis biology provided excellent markers, which allowed
655 assignment of cell stage identity with high confidence for many of these clusters. These
656 assignments provide abundant opportunities to identify new, *in vivo* stage-specific markers to test
657 for function in each stem cell lineage.

658 **Each stem cell lineage exhibited complexities, revealing known and previously**
659 **unappreciated expression dynamics**

660 In the testis, differentiation proceeds stepwise from the apical tip of the tubule. It is
661 tempting to equate this progression with the arrangement of cell states revealed in the UMAP.
662 However, the UMAP represents gene expression space and not physical space. This means that

663 a stage characterized by dramatic changes in transcription in a differentiating lineage, while
664 localized to a very small region within the organ, might be spread out in a thin stream in the UMAP.
665 An example is seen in early spermatocytes, from completion of premeiotic S phase through to
666 apolar spermatocytes (Fuller, 1993). Physically located in a narrow band next to the
667 spermatogonial region in the testes, nuclei at these stages extend across much of the lower region
668 of the germ cell UMAP (Figure 2A). A reciprocal case is highlighted by early spermatids, which
669 have very low levels of transcription and thus show very low UMI values. These cells are clustered
670 together in the UMAP, even though the cells are undergoing dramatic changes in morphology
671 easily visible by microscopy (Fuller, 1993).

672 For the soma, the complexity observed within portions of the cyst lineage was a surprise.
673 Head and tail cyst cells execute very different roles for the spermatid bundles with which they
674 associate. Not surprisingly, the distinction between head and tail cyst cells is clear late in the
675 differentiation process in the UMAP, when the two cell types have very different morphologies.
676 Our analysis confirmed that both head and tail cyst cells are derived from the same progenitor
677 population (Figure 6D). However, the somatic clusters show complex intertwining before they
678 eventually resolve to generate the two lineages. Interestingly, this complex tangling seems to
679 coincide with the stages of spermatid development where there already is clear polarization in the
680 architecture of the bundle (Figure 6D). Only by reprojecting this lineage, purposefully preserving
681 an extra dimension, could we infer the trajectory properly, discerning a split and then a merge in
682 expression profiles that tracked with cell identity. The initial split in the cyst lineage, as well as
683 the subsequent and transient merge followed by a new split, suggests an interesting sequence of
684 transcriptional cell states within this key supporting lineage.

685 **Differentiated cells map as discrete groups**

686 Whereas the two stem cell lineages each appear as strings reflecting progression in their
687 gene expression patterns, terminally differentiated cell types appear as more discrete patches in
688 the UMAP (Figure 1F). This was the case for differentiated cell types that are an integral part of
689 the testis, such as the hub and terminal epithelia, as well as for cell types that associate with the
690 testis but are structurally distinct, such as the seminal vesicle. Even within a patch, however,

691 increasing cluster resolution reveals complexity in cell type identity. For example, the snRNA-seq
692 identifies a 'muscle' group (Figure 1B, 'mc') that is composed of several different Leiden 6.0
693 clusters (Figure 2 - figure supplement 1, compare A with B). Perhaps the clusters reflect different
694 muscle cell types, for example those covering the testis tubule vs. the seminal vesicle (Susic-
695 Jung et al., 2012). Likewise, different clusters that comprise 'seminal vesicle' may represent
696 distinct portions of the structure, such as entry and exit points (Figure 8A).

697 **The value of sn- versus sc-RNA-seq comparison**

698 When carried out on the same tissue, single nucleus RNA-seq is typically comparable to
699 single cell RNA-seq, with very high percent similarity in gene identification (McLaughlin et al.,
700 2021). In the *Drosophila* testis, this was largely the case for germ cells early in the differentiation
701 lineage (Figure 4 G and H). In contrast, comparing these two approaches for later stage germ
702 cells revealed a striking difference, reflecting an important aspect of testis biology. Our data show
703 that directly comparing sn- to scRNA-seq can highlight cases where mRNAs are expressed at an
704 early developmental stage (e.g. in spermatocytes) then stored for later use (e.g. in spermatids,
705 where mRNAs may be utilized temporally). snRNA-seq likely reports mRNAs being actively
706 transcribed at a given developmental time and therefore may be more sensitive to dynamic
707 changes in transcription than scRNA-seq. Thus snRNA-seq may be a better approach for
708 mapping changes in cell state, for example, during embryonic development or differentiation in
709 stem cell lineages. Data from scRNA-seq, by contrast, may be strongly influenced by mRNAs
710 perduring in the cytoplasm from earlier time points, as well as mRNAs no longer being actively
711 synthesized but purposefully stored for later usage. Thus, while snRNA-seq reveals gene
712 expression dynamics through *active* transcriptional changes, scRNA-seq can capture post-
713 transcriptional gene regulation, as is required by transcriptionally silent cells like early spermatids.

714 Additionally, as we show here, comparison between the two approaches was key to
715 revealing programs where perdurance of mRNAs plays an important role. Cytoplasmic storage
716 of transcripts expressed at an earlier stage to be used later is especially predominant during
717 oogenesis and spermatogenesis. Many mRNAs expressed during oogenesis are stored in the
718 cytoplasm in a translationally silent state, to be recruited for translation in the early embryo

719 (Jenkins et al., 1978). This is especially important in organisms with large, yolk-rich eggs, in which
720 transcription from the zygotic genome is delayed until after several rounds of mitotic divisions. In
721 the male germ line as well, where transcription ceases during spermiogenesis as the nucleus
722 compacts, many mRNAs expressed at earlier stages are stored in the cytoplasm, initially
723 translationally repressed, then recruited for translation during spermatid morphogenesis. In both
724 cases, perdurance of the mRNAs in the cytoplasm after transcription has shut down is important
725 to allow subsequent stages of development and differentiation to take place, and recruitment of
726 different mRNAs for translation may play important roles in temporal control of morphogenetic
727 events.

728 It is also notable that just 18 of the spermatid transcribed genes encode proteins detected
729 in the mature sperm proteome (Wasbrough et al., 2010). While this might be due to limited
730 sensitivity in proteome detection, alternatively, many may play roles in spermatid development
731 but not mature sperm function. Examples could include regulating or mediating spermatid
732 elongation, the histone-to-protamine transition, or individualization, as is the case for *soti* (Barreau
733 et al., 2008; Kaplan et al., 2010). A further twenty of the 22 cytoskeletal or motor-related genes
734 transcribed in spermatids (see Figure 5 - source data 1 for functional classes of the spermatid
735 transcribed genes) have predicted functions in microtubule assembly, flagellar axoneme
736 assembly, axonemal dynein regulation or microtubule transport, consistent with roles in
737 elongating and assembling the 1.8mm sperm flagellum and transporting cargos within this large
738 cell (Ghosh-Roy et al., 2004; Noguchi et al., 2011; Tokuyasu, 1975). The set of identified lipid
739 synthesis, lipid transfer and membrane trafficking genes could contribute to membrane addition
740 at the distal (growing) ends to facilitate cell elongation (Ghosh-Roy et al., 2004). Finally, genes
741 encoding RNA binding proteins could regulate transcript localization or translation in the
742 increasingly long spermatid cells, as polarized mRNA localization by RNA binding proteins such
743 as Orb2 (localizing *aPKC* and *orb2* mRNA) and Reptin and Pontin (localizing axonemal dynein
744 mRNA) has been observed and shown to be important for sperm maturation (Fingerhut and
745 Yamashita, 2020). Whether and how many other mRNAs are localized to the growing flagellar
746 tip in spermatids remains to be studied.

747 **Use as a resource**

748 With the transcriptional profiling of over 44,000 nuclei isolated from testis and associated
749 supporting tissues, we have connected differentiation events throughout the germline and somatic
750 lineages, capitalizing on the extensive literature on *Drosophila* testis biology. The expectation is
751 that this will be a foundational resource for the field. Several other *Drosophila* RNA-seq and
752 scRNA-seq efforts have been reported (Gan et al., 2010; Hof-Michel and Bökel, 2020;
753 Mahadevaraju et al., 2021; Shi et al., 2020; Vedelek et al., 2018; Witt et al., 2019). Since each
754 approach has different relative strengths and limitations, the foundation we have laid with the
755 scRNA-seq and snRNA-seq datasets described here should assist others in comparisons with
756 more stage-restricted transcriptome analyses. More broadly, the data presented here, in their
757 easily shared formats, should enable a deeper exploration of the conserved aspects of germline
758 and support cell biology during *Drosophila* and mammalian spermatogenesis.

759

760

761 Materials and Methods

762 ***Drosophila lines***

763 For snRNA-seq, testes and attached seminal vesicles were dissected from 0-1 day old w^{1118} males
764 and processed as described in (Li et al., 2022). For scRNA-seq, testes alone were dissected
765 from 1-5 day old Oregon-R males. Oregon-R testes were also used for *in situ* hybridization of *aub*
766 and *fzo* (Figure 2). y^1w^1 flies were used for smFISH in Figure 5. Amph::GFP (CPTI-002789),
767 piwi::GFP (CPTI-003588), Akr1B::GFP (CPTI-002728), CG3902::GFP (CPTI-100001) and
768 Tup::GFP (BDSC line 81278) were from CPTI and Bac collections (Kudron et al., 2018; Lowe et
769 al., 2014). $comr^{z2-1340}$ homozygotes and w^{1118} were used for colorimetric *in situ* hybridization.

770 ***Testis squashes and analysis of expression of fluorescent fusion proteins***

771 Testes from transgenic flies of the YFP CPTI collections were dissected in testis buffer (183mM
772 KCl, 47mM NaCl, 10mM Tris pH6.8), cut open using tweezers and gently squashed on a glass

773 slide by application of a coverslip. Testis squashed preparations were imaged live in sequentially
774 captured images by phase contrast and epifluorescence microscopy using an Olympus Bx50
775 microscope, with 20x, 0.60 NA, 10x, 0.30 NA, or 40x, 0.75 NA UPlanFl objectives and either a
776 JVC KY F75U or a Hamamatsu Orca-05G camera.
777
778 For Tup::GFP analysis, testes were dissected from BAC transgenic flies, fixed in 4%
779 paraformaldehyde for twenty minutes, and blocked in 3% bovine serum albumin, 0.02% sodium
780 azide, 0.1% Triton X-100 in phosphate-buffered saline for one hour. After a one hour wash with
781 PBX (0.1% Triton X-100 in 1X phosphate-buffered saline, pH 7.4), testes were incubated
782 overnight at 4°C with antibodies to GFP (Abcam 13970; 1:10,000) and Fasciclin-III
783 (Developmental Studies Hybridoma Bank 7G10; 1:50). After washing with PBX testes were
784 treated with goat anti-chicken AlexaFluor 488 (Invitrogen A, 11039; 1:200), goat anti-mouse
785 AlexaFluor 568 (Invitrogen A11004; 1:100), and DAPI (Millipore Sigma 1023627600; 1:1,000) for
786 two hours. After final washes in PBX, testes were mounted in VectaShield (Vector Labs H-1000),
787 and images were captured at 63X, NA 1.4, on a Zeiss LSM800 confocal microscope.

788 ***RNA in situ Hybridization***

789 **For *in situs* presented in Figures 2 and 4:**
790 *aub* (forward primer 5'-CCTGGGCGGCTACATCTT-3'; reverse primer 5'-
791 GCGCAGATTTCGACTCGG'-3), *kmg (*forward primer 5'-TGCCTCTATGCCTCACGC-3'; reverse
792 primer 5'- GCGCCTACCGGTCTCATC-3''*), fzo* (forward primer 5'-GGCATCCAAACTCTCGCG-
793 3';
794 reverse primer 5'-TGTCGCAACTGGAGCTCA-3') were amplified by PCR on cDNA from Oregon-
795 R Drosophila Testes. Using TA cloning (Promega, "Easy-T" Cloning), the resulting amplicons
796 were cloned into th pGEM vector (Promega). Subsequent PCR added a T7 binding site (5'-
797 GAAGTAATACGACTCACTATAGGGAGAGGG-3') upstream of the amplicon. The resulting
798 plasmids then served as templates for *in vitro* transcription with Digoxigenin (DIG)- and
799 fluorescein isothiocynate (FITC)-labeled ribonucleotides to generate labeled single-stranded
800 antisense riboprobes.

801

802 Testes were isolated and fixed in 4% formaldehyde for 30-60 minutes, dehydrated into methanol
803 and stored at -20°C for up to 1 month. After rehydration in PBS + 0.1% Triton-X100, testes were
804 permeabilized with 4 ug/ml Proteinase K for 6 minutes and washed with Pre-Hybridization solution
805 (50% deionized formamide, 5x SSC, 1mg/ml yeast RNA, 1% Tween-20) for up to 2 hours at 56°C.
806 Testes were incubated overnight at 56°C with probes diluted 1:800 in Hybridization solution (50%
807 deionized formamide, 5x SSC, 1 mg/mL yeast RNA, 1% Tween-20, and 5% Dextran Sulfate).
808 After washes in Pre-Hybridization solution, 2x SSC, and 0.2x SSC, then PBS+0.1% Triton-X100,
809 samples were blocked for 30 minutes in 1% Roche Western Blocking Reagent prior to incubating
810 overnight at 4°C with either anti-FITC with horseradish peroxidase conjugate (Roche) at a 1:2000
811 concentration or anti-DIG- with horseradish peroxidase conjugate (Roche) at a 1:1500
812 concentration in 1% Roche Western Blocking Reagent. Fluorescent tyramide development and
813 amplification were performed by first placing the testes for 5 min in borate buffer (0.1M boric acid,
814 2M NaCl, pH 8.5), followed by 10 min in borate buffer with rhodamine (1:1000) or fluorescein
815 (1:1500) tyramide, and 0.0003% hydrogen peroxide. After development, peroxidase activation
816 was performed in a 1% sodium azide solution for at least 1h, followed by antibody labeling for the
817 second probe. Coverslips were mounted in Vectashield with DAPI (Vector Labs). Fluorescence
818 image acquisition was performed on a Leica Stellaris Confocal microscope using a 63X oil-
819 immersion objective (NA = 1.4).

820

821 **For *in situs* presented in Figures 5, 6, and 8:**

822 Fixation and hybridization for FISH was as described for Figures 2 and 4, with minor modifications
823 as described (Wilk et al., 2017). Briefly, these include using a cold acetone permeabilization step
824 and 0.3% Triton X-100 instead of Proteinase K for improved tissue permeabilization, and DIG-
825 labeled probe detection via tyramide amplification. DAPI (4" ,6-diamidino-2-phenylindole; Sigma,
826 cat. no. D-9542) was used at 1 ug/ml to reveal nuclei. For detection of *piwi*, *eya*, *f-cup*, *loqs* and
827 *geko* transcripts, RNA probes were transcribed from the BDGP DGC library plasmid clones
828 GM05853, GH05272, GH09045, RE14437 and RE30284, respectively using T7 RNA

829 polymerase. For *hh* and *Nep5*, templates were made by PCR from genomic DNA using the
830 following T7 and T3 promoter-containing primers:
831
832 *Hh* last exon:
833 Forward: GTAATACGACTCACTATAGGGAGACCACTGCCGATTGATTTTCTCAGG
834 Reverse: AATTAACCCTCACTAAAGGGTTGTGGAGATCGTGTTTTGAGCAT
835
836 *Nep5* exon 6:
837 Forward: GTAATACGACTCACTATAGGGAGACCACGGGGAAATCCGATAAAGCTC
838 Reverse: AATTAACCCTCACTAAAGGGTTGTATCTGCAGAACCAAACTGAC
839
840 **For colorimetric *in situ* hybridisation in Figure 3, 8 and Figure 3 - figure supplement 1**
841 Probe preparation and *in situ* hybridisation were performed as described (Morris et al., 2009).
842 Primers Hml-F ATTTAGGTGACACTATAGAATAAGTGGACCCATGCCAAG and Hml-R
843 TAACCCTCACTAAAGGGTGACCATCATCGCAAATC and. primers MtnA-F
844 ATTTAGGTGACACTATAGAAGCGGTAAGTTCGCAGTC and MtnA-R
845 TAACCCTCACTAAAGGGACATTTATTGCAGGGTGTG were used to amplify 628bp and 443bp
846 fragments respectively from cDNA generated from w^{1118} testes. After re-amplification using
847 primers 5'-SP6 ACGGCAATTTAGGTGACACTATAGAA and 3'-T3
848 GCAACGAATTAACCCTCACTAAAGGG, and the products served as templates for T3 RNA
849 polymerase to generate dig-labeled single stranded, antisense RNA probes. The probes were
850 hydrolysed for 15 minutes, precipitated and resuspended in 200 µl water. Testes were dissected
851 from young w^{1118} males, fixed in 4% paraformaldehyde for 20-60 minutes, washed in PBS,
852 permeabilised with 50 µg/ml proteinase K for 5 minutes, washed in PBS, then hybridisation buffer
853 (HB: 50% Formamide, 5× SSC, 100 µg/ml denatured sonicated salmon sperm DNA, 50 µg/ml
854 heparin, 0.1% Tween 20, 100 mM citric acid). Probes were diluted 1:100 in HB and testes were
855 hybridized for 16 hr at 65°C. Testes were washed 6x30 min at 65°C in HB, followed by 15 min
856 each step at room temp in 4:1, 3:2, 2:3 1:4 HB:PBST, 2x15 min PBST, then incubated overnight
857 at 4°C in alkaline phosphatase-conjugated anti-digoxygenin antibody diluted 1:2,000 in PBST.

858 Testes were washed 4x15 min in PBST and finally 3x 5 min HP (100 mM NaCl, 100 mM Tris (pH
859 9.5), 50 mM MgCl2, 0.1% Tween 20). NBT and BCIP diluted in HP were added as a colorimetric
860 substrate for alkaline phosphatase and color allowed to develop in the dark at room temperature.
861 The testes were washed in PBST, dehydrated through an ethanol series, mounted in GMM, and
862 imaged using DIC microscopy (10x objective) with a JVC KY F75U camera mounted on an
863 Olympus BX50 microscope.
864
865 **For single molecule FISH presented in Figure 5**
866 smFISH was performed as previously described (Fingerhut et al., 2019). Briefly, testes were
867 dissected in 1xPBS (Invitrogen, AM9624) and fixed in 4% formaldehyde (Polysciences, Inc.,
868 18814-10) in 1xPBS for 30 minutes, washed twice in 1xPBS for 5 minutes each, and
869 permeabilized in 70% ethanol overnight at 4°C. Testes were washed with wash buffer (2x saline-
870 sodium citrate [SSC, Invitrogen, AM9770], 10% formamide [Fisher Scientific, BP227]), and then
871 hybridized overnight at 37°C in hybridization buffer (2xSSC, 10% dextran sulfate [Sigma-Aldrich,
872 D8906], 1mg/mL yeast tRNA [Sigma-Aldrich, R8759], 2mM Vanadyl Ribonucleoside complex
873 [New England Biolabs, S1402S], 0.5% BSA [Invitrogen, Am2616], 10% formamide). Following
874 hybridization, samples were washed twice with wash buffer for 30 minutes each at 37°C and
875 mounted in Vectashield with DAPI (Vector Laboratory, H-1200, Burlingame, USA). Images were
876 acquired using an upright Leica Stellaris 8 confocal microscope with a 63X oil immersion objective
877 lens (NA = 1.4) and processed using ImageJ and Adobe Photoshop software (National Institutes
878 of Health, Bethesda, USA).
879
880 Fluorescently labeled probes were added to the hybridization buffer to a final concentration of
881 100nM. Probes against *soti* and *wa-cup* transcripts were designed using the Stellaris® RNA FISH
882 Probe Designer (Biosearch Technologies, Inc., Novato, USA) available online at
883 www.biosearchtech.com/stellarisdesigner.

Probe Target	Fluorophore	5'-Sequence-3'

wa-cup	Quasar® 670	cttccgataagctcatttgg,	ccgctgttgggtgaaaaaga,
		ccagagcgctgttgaaatac,	gagctctttcattgaacgga,
		atcgatcttttcagctgact,	ctttcggctcatcaacagat,
		gactgacattgggattggtt,	tacagagcatcgcagacttc,
		gaatcttccaagcgattgga,	ccggagctaaatcgctttaa,
		tgagacggtcgagaacagga,	gacatggtggtatcatctga,
		aactagccatcatgcgattg,	gtatccttaatatccttggg,
		gttttccattatgctaacca,	agtggggacaaatgggtttc,
		ttaaaatgccttttcgcct,	ccagcatttgttcagatacg,
		tcgcagtgtcttcagaaagg,	gactttttgcaatgcttggg,
		catcagtttgcccaaatact,	tgcgttccgctttataattg,
		taaggcgtacatgggacttc,	ctcgttgggatatttctgtt,
		ctcccggtgcttattataaa,	gtctggtactgaatgcgata,
		atgctgcgcagaatcttgaa,	gttctccaactcgaattagc,
		cccatgacttcctcaataaa,	cgtcttgatggtgacatagt,
		aaatttccacggcattacgc,	aatgcgagctaaacccaagt,
		agacagtcatattgctggga,	gtgttagcagacgttgtttg,
		actcgttcgttttgtctttg,	gtagccgatctggttatatt,
		ttaaaatgctccgctttggg,	atacgattttccagtcggac,
		gttcaatgtgatactcggca,	aattcgtgcagtagataggc,
		acaattcagatgctcttggg,	ctccatataacactcttgca,
		gccttgcatataaccatgag,	taagcacaggtcaaggttct,
		ctcctccgcattaactttaa,	tgagccaaacttttgtctct,
		ctgatcgttgctttggaaca,	taatttggttgcgatcctca

soti	Quasar® 570	ctcgacgaggtaatttg, gctcatcgtacagatcgt, atcttcattcaccgcgtc, tgctgtccatcctccaat, gtccaggagtatgtccat, ctccttgcgccggaaaaa, aacttcgtttcttccgcc, ctcctgactttggcatgg, attgccctcgtgacactg, caaagtactcgcctcgct, tggcagacccataccatt,	tccgtgtagtacgtccat, ccgactcgatcgattagc, tgtccaagtcatcgccag, tgacgattgactcccagg, caacggtggctcttgagg, acgtggtggtccatttgg, ggagtgggtttggtcata, ttaggaggcacatctccg, atcctcgcgaacgtgacg, gggtagttctgactggtc, agaactgaccccaatgct

Preparation of samples for snRNA-seq

FACS-sorted nuclei were obtained from hand dissected, 0-1 day-old adult testes (plus seminal vesicles) and processed as described by the Fly Cell Atlas project (Li et al., 2022). Data from 44,621 sequenced nuclei passed quality control metrics. The raw data are publicly available (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-10519/).

Importing FCA data into Seurat

The “Testis, 10X, Relaxed” loom file (https://flycellatlas.org/; Li et al., 2022) was imported into Seurat 4.0.5, and a standard pipeline run on the resulting Seurat R object to normalize and scale the data (NormalizeData, FindVariableFeatures, and ScaleData). The loom file had already been filtered with nuclei expressing less than 200 genes or exceeding 15% mitochondrial content removed and genes not expressed in at least 3 nuclei removed. We chose to use the relaxed rather than a ‘10X, Stringent’ testis loom file due to the particular biology of the testis. Testis germline cells can express ‘somatic’ genes (see Results), including the hub cell marker *upd1*. As a consequence, filtering algorithms that generated the stringent dataset led to loss of a documented somatic cell type (the hub), as well as a large number of late spermatocytes. The Fly Cell Atlas website provides links and tutorials to ASAP and SCope, two web-based pipeline

901 and visualization portals where users can examine the data or re-run analyses (Davie, 2018;
902 Gardeux, 2017). Within the SCope interface, select the 10X > Testis > Relaxed dataset, and
903 Settings > HVG UMAP. The analysis in this paper complements SCope and ASAP with Seurat
904 and Monocle3, two R programming-based tools for single cell analysis (Satija et al., 2015;
905 Trapnell et al., 2014 , Tools for Single Cell Genomics • Seurat; https://cole-trapnell-
906 lab.github.io/monocle3/).
907
908 The cluster information contained in the original loom file was preserved in the Seurat object, with
909 clusterings available at increasing levels of resolution (Leiden algorithm, 0.4 to 10). The level of
910 granularity provided by resolution 6.0 was deemed most revealing and is used for most analysis
911 here. The full Seurat Object, FCAloomToSeurat2TFP_Annotations.rds, is available for download
912 at Input Files.
913
914 Original cluster 90 was manually split into two in the Seurat Object,
915 FCAloomToSeurat2TFP_Annotations.rds, with one resultant cluster of 79 nuclei definitively
916 established as representing the hub (retaining its cluster number, #90), and the remaining nuclei
917 placed into a new cluster, #111. Venn diagrams were created using Venny 2.1 (Oliveros, J.C.
918 2007-2015).
919
920 In some cases, data were extracted from subsets of the original Seurat object: a germline only
921 subset (Figures 2 through 5), a somatic cyst cell lineage subset (Figure 6), and a subset
922 representing several epithelial cell types of the testis as well as specific additional epithelial cell
923 types (Figure 8). All UMAPs were generated within Seurat using the “DimPlot” function. UMAPs
924 that highlighted particular genes of interest were generated using the “FeaturePlot” function
925 (Figures 1, 2, 3, 5, 6, 7, 8). Dotplots were generated using the “DotPlot” function (Figures 2
926 through 7).
927
928 The testis 10X Relaxed loom file only contains a UMAP reduction projected to two dimensions.
929 To inspect a 3D UMAP representation, the appropriate lineage was isolated and reprojected,

930 passing the argument 'n.components =3L' to the Seurat function 'RunUMAP' (see
931 cystlineage3Dcode.R code; (Qadir et al., 2020); 10.5281/zenodo.348317). An html version of the
932 resulting 3D representation is available for download at Input Files.
933
934 For analysis of heart, Malpghian tubule and male reproductive gland (Figure 3 - figure supplement
935 1), the appropriate stringent 10X loom file from Fly Cell Atlas was imported into Seurat. Means
936 were calculated for nCount (UMI) and nFeature (gene). Cluster numbers were assigned using
937 the "Annotation" metadata field in each object, and a plot produced for average UMI by cluster
938 number, with dot size reflecting average gene number per cluster.
939
940 Graphs generated directly from ASAP were used to produce Figures 3D-E (Gardeux *et al.*, 2017).
941 Continuous Coloring of a Custom Gene Set - Categorical Gene Metadata: _Chromosomes (either
942 X or Y) was used under the Visualization tab. SVGs can be saved directly from the website. The
943 enrichment analysis performed to produce these graphs mirrors Seurat's AddModuleScore
944 function and is detailed in the Materials and Methods of Li et al., 2022 (see header: Metabolic
945 clustering using ASAP).
946
947 On occasion, use of an alternative Assay ("log.counts") in the Seurat Object allowed for
948 visualization of low levels of gene expression in spermatocytes. This "log.counts" assay contains
949 a matrix of $\log_2$(counts + 1), or log-transformed raw counts. This was done for plotting or to
950 perform analyses focused on promiscuous expression in spermatocytes (Figures 3F-G, 3J, Figure
951 3 - figure supplement 1D-G, Figure 3 - source data 1). Figure 3 code describes how this assay
952 was added and shows how to utilize this information when needed.
953
954 Cell type-specific transcription factors were taken from Li et al., 2022, Table S3 and are available
955 at Input Files (TFs_list_500.txt). Each gene was classified as being only in mature spermatocytes,
956 both in other cell types and mature spermatocytes, or not in germ line cell types as per
957 assignments in the Fly Cell Atlas Table S3 (Li et al., 2022). Figure 3 - source data 1 was generated
958 by calculating average log-transformed expression of each transcription factor per Leiden 0.4

959 cluster. Subtracting the value in cluster 3 from that of cluster 2 yielded the upregulation of each
960 gene in mature spermatocytes relative to spermatogonia and early spermatocytes (Figure 3J,
961 Figure 3 - source data 1).

962 ***Tissue isolation and cell dissociation for scRNA-seq***

963 The testis dissociation protocol was adapted from Witt et al., 2019. Fresh maceration buffer
964 (10mL Trypsin LE (Gibco) with 20 mg collagenase (Gibco)) was prepared on the day of the
965 dissection. Testes were hand dissected from 1-5 day old male flies in 1x phosphate-buffered
966 saline (PBS), separated from seminal vesicles and transferred immediately into tubes filled with
967 cold PBS, on ice. Testes were kept in PBS for a maximum of 30 minutes. Samples were
968 centrifuged at 135 rcf and the PBS removed and replaced with 400 µL marceration buffer. Testes
969 were incubated in maceration buffer for 30 minutes with gentle vortexing every 10 minutes at
970 room temperature. Following incubation, samples were pipetted up and down for 15 minutes until
971 all visible chunks were gone and the sample was in approximately a single-cell suspension.
972 Sample was filtered through a 35 µm filter into a polystyrene tube, then transferred into a
973 microcentrifuge tube. After the sample was centrifuged at 135 rcf for 7 minutes, the supernatant
974 was removed and the pellet resuspended in 1mL calcium- and magnesium-free Hanks' Balanced
975 Salt Solution (HBSS). The sample was spun a final time at 137 rcf for 7 minutes. All but 50 µL
976 of the HBSS supernatant was removed, and the cell pellet was resuspended in the remaining 50
977 µL. Cell viability and density was then assayed on a hemocytometer using DIC imaging and
978 Trypan Blue stain.
979

980 ***Library preparation and sequencing for scRNA-seq***

981 Cells were processed using the 10x Genomics Chromium Controller and Chromium Single Cell
982 Library and Gel Bead Kit following standard manufacturer's protocol. Amplified cDNA libraries
983 were quantified by bioanalyzer and size selected using AMPure beads. Samples were sequenced
984 on a NovaSeq SP.

985 ***Mapping and preprocessing of scRNA-seq data***

986 Reads were mapped to the DM3 reference genome using the 10X CellRanger pipeline with
987 default parameters. The resulting feature matrix (default output, kept in
988 outs/filtered_feature_bc_matrix, and featuring barcodes.tsv.gz, features.tsv.gz, and
989 matrix.mtx.gz) was read into the R package Monocle3 using load_cellranger_data. The resulting
990 cell data set (cds) object was processed using 100 dimensions, and underwent dimensionality
991 reduction using the UMAP method. Germline cells were identified on the basis of super-cluster
992 (in Monocle3, "partition") identity, with 100 dimensions used to identify partitions. The germline
993 cells were then subsequently clustered using the Monocle3 "cluster_cells" command, with
994 resolution = 0.003.

995 ***Trajectory analysis of FCA snRNA-seq data***

996 The publicly available "Testis, 10X, Relaxed" loom file of snRNA-seq data from the Fly Cell Atlas
997 website (https://flycellatlas.org/; 10.1101/2021.07.04.451050) was read into Monocle3, and
998 preprocessed using 50 dimensions. This dimension number was determined empirically as it
999 resulted in connected clusters that represented the primary lineages (germline and cyst). UMAP
1000 dimensionality reduction and clustering was performed with a resolution of 0.0002, again
1001 determined empirically to represent biologically significant clusters that approximated the original
1002 annotation. A trajectory graph was generated from data with "learn_graph". Pseudotime was
1003 calculated with "order_cells", with the first (base) node selected as the root in Monocle3's
1004 interactive mode. Pseudotime parameters were then subsequently visualized on the original
1005 projection by adding a "pseudotime" slot to the FCA Seurat object. "learn_graph" and "order_cells"
1006 were likewise run on the scRNA-seq dataset, with the single most base node again selected as
1007 the root.
1008
1009 A list of genes that change expression level dynamically along pseudotime was generated from
1010 the full join of genes that vary along pseudotime in the single cell and nucleus datasets, according
1011 to the graph_test function in Monocle3 (parameters: neighbor_graph = "principal_graph", method
1012 = "Moran_I" cores = 4; selected genes with q_value = 0 and morans_I = 0.25). Additionally,

1013 several genes encoding transcripts that were annotated as enriched in late pseudotime in the
1014 original FCA paper were added to the analysis. Genes (representing rows on the Figure 4G-H
1015 heatmaps) were ordered according to pseudotime point of peak expression averaged between
1016 the two datasets. A z-score for each gene expression for each dataset, smoothed across
1017 pseudotime using R function "smooth.spline" with 3 degrees of freedom was calculated and the
1018 heatmap was generated using the R package ComplexHeatmap.

1019 ***Aligning trajectories for scRNA-seq and FCA snRNA-seq data by dynamic time warping***

1020 The trajectories for the scRNA-seq and snRNA-seq data were aligned on a common "warped"
1021 time scale using a Dynamic Time Warping based procedure (adapted from Alpert et al., 2018;
1022 Cacchiarelli et al., 2018). The smoothened gene expression (as shown in the heatmaps) for all
1023 germline cells in each of the two monocle3 trajectories was used for alignment using the dtw
1024 function in R package dtw. The Pearson's correlation based distance and "symmetric2" step
1025 parameters were used in dtw. The aligned time scale returned by dtw was used as the warped
1026 pseudotime. Transcript expression levels could then be plotted on the same axis, each as a
1027 proportion of its own pseudotime.
1028

1029 ***Code and Data Availability***

1030 Scripts, written in the programming languages R or Python, necessary to reproduce the analyses
1031 in this manuscript are contained in a folder hosted at: Code. Data files (.loom, .txt, .csv., .tsv, or
1032 .rds) necessary as input to the above scripts are stored at Input Files. Both the raw and the
1033 filtered data from the FCA snRNA-seq analysis are publically available as indicated in the Results
1034 section of the text. The scRNA-seq data and Moncle object are available at
1035 https://drive.google.com/drive/folders/1A2U5piOy4-HDLloFezux2dZOCGxznDQS?usp=sharing.

1036 **Acknowledgements**

1037 We especially acknowledge and thank the Fly Cell Atlas project, the Chan-Zuckerberg Biohub,
1038 the NIH sequencing facility and the laboratory of Steve Quake for the starting testis plus seminal

vesicle snRNA-seq dataset. We thank Stein Aerts and Bart Deplancke for supporting the efforts of J.J. and W.S., respectively. We acknowledge the Whitehead Institute Genome Technology Core for assistance in library preparation and sequencing of the scRNA-seq dataset, the Bloomington Drosophila Stock Center and Kyoto DGGR for *Drosophila* strains, the Developmental Studies Hybridoma Bank for various antibodies, and Aaron Allen for advice and initial code to transfer the SCope loom file format into SEURAT. S.D. thanks Dan Beiting at PennVet for his DIY_Transcriptomics course.

Funding

NIH F32 Award number 1F32GM143850-01 (Amelie Raz); Howard Hughes Medical Institute (Yukiko Yamashita); CIHR Project grant PJT-133743 (Henry Krause and Julie A. Brill); NSERC Discovery grants RGPIN-2016-06775 and RGPIN-2022-05163 (Julie A. Brill); NIH R35GM136665 and R01HD052937 (Erika Matunis); NSF 2021316576 (Jenn Viveiros); NSF 2022305201 (Jasmine Grey); NIH F31 1F31HD108918-01 (Mara Grace); NIH R35GM136433 and the Reed-Hodgson Professorship in Human Biology (Margaret Fuller); CPRIT Scholar in Cancer Research RR200063 and NIH R00AG062746 (Hongjie Li); NIH R35GM136270 (Stephen DiNardo); NIH T32 GM008216 (Gabriela Vida); NIH T32 GM007790 and a Stanford Graduate Fellowship (Sarah Stern); NIH T32 T32AR007422 (Cameron Berry); Biotechnology and Biological Sciences Research Council BB/P001564/1 (Helen White-Cooper); Brian Oliver was supported in part by the Intramural Research Program of the National Institutes of Health (NIDDK).

Competing Interests

No competing interests declared

Author Contributions

AAR, GSV, SRS, SM, JMF, JMV, SP, JRG, and MRG ('contributed equally') each took the lead on a particular aspect of the coding, analysis, documentation, construction of at least one specific

Figure and or its supplement, and/or assembly of supporting source data, and also re-interatively drafted sections of the narrative reporting that work. CWB led initial analysis of the testis snRNA-seq dataset and was involved in justifying the progression along the lineages. HL prepared and processed and initially analyzed the snRNA-seq samples of testis plus seminal vesicle and AAR prepared and processed and analyzed the scRNA-seq samples. JJ aided the transition from analysis in SCope to SEURAT and WAS assisted with the trajectory analyses. ZS and CH (FlyFish) designed probes and carried out *in situ* hybridizations, imaging and analysis. YY helped develop, oversee and advise on the scRNA-seq carried out by AAR. YY provided substantive guidance and advice in discussions with AAR and JF, and TP did so with SP. BO, JAB, HMK, ELM, HW-C, SD, and MTF each contributed framing of key questions, data, data interpretation, insight, and knowledge in drafting various sections of the manuscript, and/or provided images and/or code to interrogate the points considered. Virtually all authors made substantive comments impacting the direction of the work during many remote jamboree meetings, as well as commenting insightfully on the shared manuscript and Figure drafts. SD provided guidance on R and assisted with data analysis in Seurat. MTF recognized the need to communicate more than had been possible in the original FCA paper, organized the collaboration to do so, outlined figures, and wrote the first draft. Both SD and MTF helped oversee the efforts of the team, and contributed significantly to manuscript editing.

1084 **References**

1085 Alpert, A., Moore, L.S., Dubovik, T., and Shen-Orr, S.S. (2018). Alignment of single-cell
1086 trajectories to compare cellular expression dynamics. Nat. Methods *15*, 267–270.

1087 Alphey, L., Jimenez, J., White-Cooper, H., Dawson, I., Nurse, P., and Glover, D.M. (1992).
1088 twine, a cdc25 homolog that functions in the male and female germline of Drosophila. Cell *69*,
1089 977–988.

1090 Amoyel, M., Sanny, J., Burel, M., and Bach, E.A. (2013). Hedgehog is required for CySC self-
1091 renewal but does not contribute to the GSC niche in the Drosophila testis. Development *140*,
1092 56–65.

1093 Anllo, L., and DiNardo, S. (2022). Visceral mesoderm signaling regulates assembly position and
1094 function of the Drosophila testis niche. BioRxiv 2021.09.08.459436.

1095 Anllo, L., Plasschaert, L.W., Sui, J., and DiNardo, S. (2019). Live imaging reveals hub cell
1096 assembly and compaction dynamics during morphogenesis of the Drosophila testis niche. Dev.
1097 Biol. *446*, 102–118.

1098 Barckmann, B., Chen, X., Kaiser, S., Jayaramaiah-Raja, S., Rathke, C., Dottermusch-Heidel,
1099 C., Fuller, M.T., and Renkawitz-Pohl, R. (2013). Three levels of regulation lead to protamine and
1100 Mst77F expression in Drosophila. Dev. Biol. *377*, 33–45.

1101 Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F., and White-Cooper, H. (2008). Post-
1102 meiotic transcription in Drosophila testes. Development *135*, 1897–1902.

1103 Boukhatmi, H., Frendo, J.L., Enriquez, J., Crozatier, M., Dubois, L., and Vincent, A. (2012).
1104 Tup/Islet1 integrates time and position to specify muscle identity in Drosophila. Development
1105 *139*, 3572–3582.

1106 Boyle, M., Wong, C., Rocha, M., and Jones, D.L. (2007). Decline in Self-Renewal Factors
1107 Contributes to Aging of the Stem Cell Niche in the Drosophila Testis. Cell Stem Cell *1*, 470–478.

1108 Le Bras, S., and Van Doren, M. (2006). Development of the male germline stem cell niche in
1109 Drosophila. Dev. Biol. *294*, 92–103.

1110 Brower, D.L., Smith, R.J., and Wilcox, M. (1981). Differentiation within the gonads of Drosophila
1111 revealed by immunofluorescence. J. Embryol. Exp. Morphol. *63*, 233–242.

1112 Cacchiarelli, D., Qui, X., Mikkelsent, T.S., Rinn, J.L., and Trapnell, C. (2018). Aligning Single-
1113 Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of
1114 Myogenic Reprogramming Outcome. Cell Syst. 258–268.

1115 Cao, C., Ma, Q., Mo, S., Shu, G., Liu, Q., Ye, J., and Gui, Y. (2021). Single-Cell RNA
1116 Sequencing Defines the Regulation of Spermatogenesis by Sertoli-Cell Androgen Signaling .
1117 Front. Cell Dev. Biol. *9*.

1118 Cao, J., Packer, J., Ramani, V., Cusanovich, D., Chau, H., Riza, D., Xiaojie, Q., Choli, L., N.,
1119 F.S., J., S.F., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular
1120 organism. Science *357*, 661–667.

1121 Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S.,
1122 Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of
1123 mammalian organogenesis. Nature *566*, 496–502.

1124 Chandley, A.C., and Bateman, A.J. (1962). Timing of Spermatogenesis in Drosophila
1125 melanogaster using Tritiated Thymidine. Nature *193*, 299–300.

1126 Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., Wang, Q., Xie, N., Hua, R., Liu, M., et
1127 al. (2018). Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse
1128 spermatogenesis. Cell Res. *28*, 879–896.

1129 DeFalco, T., Camara, N., Le Bras, S., and Van Doren, M. (2008). Nonautonomous Sex
1130 Determination Controls Sexually Dimorphic Development of the Drosophila Gonad. Dev. Cell
1131 *14*, 275–286.

1132 DiNardo, S., Okegbe, T., Wingert, L., Freilich, S., and Terry, N. (2011). lines and bowl affect the
1133 specification of cyst stem cells and niche cells in the Drosophila testis. Development *138*, 1687–
1134 1696.

1135 Edgar, B.A., and O'Farrell, P.H. (1990). The three postblastoderm cell cycles of Drosophila
1136 embryogenesis are regulated in G2 by string. Cell *62*, 469–480.

1137 Fabrizio, J.J., Boyle, M., and DiNardo, S. (2003). A somatic role for eyes absent (eya) and sine
1138 oculis (so) in drosophila spermatocyte development. Dev. Biol. *258*, 117–128.

1139 Fairchild, M.J., Smendziuk, C.M., and Tanentzapf, G. (2015). A somatic permeability barrier
1140 around the germline is essential for Drosophila spermatogenesis. Development *142*, 268–281.

1141 Fincher, C.T., Wurtzel, O., de Hoog, T., Kravarik, K.M., and Reddien, P.W. (2018). Cell type
1142 transcriptome atlas for the planarian Schmidtea mediterranea. Science *1736*, 1–19.

1143 Fingerhut, J.M., and Yamashita, Y.M. (2020). mRNA localization mediates maturation of
1144 cytoplasmic cilia in Drosophila spermatogenesis. J. Cell Biol. *219*, e202003084.

1145 Fingerhut, J.M., Moran, J. V., and Yamashita, Y.M. (2019). Satellite DNA-containing gigantic
1146 introns in a unique gene expression program during Drosophila spermatogenesis. PLoS Genet.
1147 *15*, 1–23.

1148 Fuller, M.T. (1993). Spermatogenesis, in The Development of Drosophila (Cold Spring Harbor,
1149 NY: Cold Spring Harbor Laboratory Press).

1150 Fuller, M.T. (1998). Genetic control of cell proliferation and differentiation in Drosophila
1151 spermatogenesis. Semin. Cell Dev. Biol. *9*, 433–444.

1152 Fuse, N., Hirose, S., and Hayashi, S. (1994). Diploidy of Drosophila imaginal cells is maintained
1153 by a transcriptional repressor encoded by escargot. Genes Dev. *8*, 2270–2281.

1154 Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K., and Chen, X. (2010). Dynamic
1155 regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by
1156 RNA-seq. Cell Res. *20*, 763–783.

1157 Ghosh-Roy, A., Kulkarni, M., Kumar, V., Shirolikar, S., and Ray, K. (2004). Cytoplasmic Dynein–
1158 Dynactin Complex Is Required for Spermatid Growth but Not Axoneme Assembly in Drosophila.
1159 Mol. Biol. Cell *15*, 2470–2483.

1160 Gönczy, P., Viswanathan, S., and DiNardo, S. (1992). Probing spermatogenesis in Drosophila
1161 with P-element enhancer detectors. Development *114*, 89–98.

1162 Gonzalez, J., Qi, H., Liu, N., and Lin, H. (2015). Piwi Is a Key Regulator of Both Somatic and
1163 Germline Stem Cells in the Drosophila Testis. Cell Rep. *12*, 150–161.

1164 Gould-Somero, M., and Holland, L. (1974). The timing of RNA synthesis for spermiogenesis in
1165 organ cultures ofDrosophila melanogaster testes. Wilhelm Roux. Arch. Entwickl. Mech. Org.
1166 *174*, 133–148.

1167 Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C.,
1168 Gurczynski, S.J., Moore, B.B., et al. (2018). A Comprehensive Roadmap of Murine
1169 Spermatogenesis Defined by Single-Cell RNA-Seq. Dev. Cell *46*, 651-667.e10.

1170 Greenspan, L.J., de Cuevas, M., Le, K.H., Viveiros, J.M., and Matunis, E.L. (2022). Activation of
1171 the EGFR/MAPK pathway drives transdifferentiation of quiescent niche cells to stem cells in the
1172 Drosophila testis niche. Elife *11*, e70810.

1173 Guo, J., Nie, X., Giebler, M., Mlcochova, H., Wang, Y., Grow, E.J., Kim, R., Tharmalingam, M.,
1174 Matilionyte, G., Lindskog, C., et al. (2020). The Dynamic Transcriptional Cell Atlas of Testis
1175 Development during Human Puberty. Cell Stem Cell *26*, 262-276.e4.

1176 Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey,
1177 T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium.
1178 Nature *551*, 333–339.

1179 Hardy, R.W., Tokuyasu, K.T., Lindsley, D.L., and Garavito, M. (1979). The germinal proliferation
1180 center in the testis of Drosophila melanogaster. J. Ultrastruct. Res. *69*, 180–190.

1181 Herrera, S.C., Sainz de la Maza, D., Grmai, L., Margolis, S., Plessel, R., Burel, M., O'Connor,
1182 M., Amoyel, M., and Bach, E.A. (2021). Proliferative stem cells maintain quiescence of their
1183 niche by secreting the Activin inhibitor Follistatin. Dev. Cell *56*, 2284-2294.e6.

1184 Hétié, P., de Cuevas, M., and Matunis, E. (2014). Conversion of Quiescent Niche Cells to
1185 Somatic Stem Cells Causes Ectopic Niche Formation in the Drosophila Testis. Cell Rep. *7*,
1186 715–721.

1187 Hof-Michel, S., and Bökel, C. (2020). Transcriptome analysis of somatic cell populations in the
1188 Drosophila testis links metabolism and stemness. BioRxiv 2020.02.24.962639.

1189 Hwa, J.J., Hiller, M.A., Fuller, M.T., and Santel, A. (2002). Differential expression of the
1190 Drosophila mitofusin genes fuzzy onions (fzo) and dmfn. Mech. Dev. *116*, 213–216.

1191 Hwa, J.J., Zhu, A.J., Hiller, M.A., Kon, C.Y., Fuller, M.T., and Santel, A. (2004). Germ-line
1192 specific variants of components of the mitochondrial outer membrane import machinery in
1193 Drosophila. FEBS Lett. *572*, 141–146.

1194 Jenkins, N.A., Kaumeyer, J.F., Young, E.M., and Raff, R.A. (1978). A test for masked message:
1195 The template activity of messenger ribonucleoprotein particles isolated from sea urchin eggs.
1196 Dev. Biol. *63*, 279–298.

1197 Kaplan, Y., Gibbs-Bar, L., Kalifa, Y., Feinstein-Rotkopf, Y., and Arama, E. (2010). Gradients of a
1198 Ubiquitin E3 Ligase Inhibitor and a Caspase Inhibitor Determine Differentiation or Death in
1199 Spermatids. Dev. Cell *19*, 160–173.

1200 Kapoor, T., Dubey, P., Shirolikar, S., and Ray, K. (2021). An actomyosin clamp assembled by
1201 the Amphiphysin-Rho1-Dia/DAAM-Rok pathway reinforces somatic cell membrane folded
1202 around spermatid heads. Cell Rep. *34*, 108918.

1203 Kawase, E., Wong, M.D., Ding, B.C., and Xie, T. (2004). Gbb/Bmp signaling is essential for
1204 maintaining germline stem cells and for repressing bam transcription in the Drosophila testis.
1205 Development *131*, 1365–1375.

1206 Kiger, A., Leanne, J.D., Cordula, S., Rogers, M., and Fuller, M. (2001). Stem Cell Self-Renewal
1207 Specified by JAK-STAT Activation in Response to a Support Cell Cue. Science *294*, 2542–
1208 2545.

1209 Kudron, M.M., Victorsen, A., Gevirtzman, L., Hillier, L.W., Fisher, W.W., Vafeados, D., Kirkey,
1210 M., Hammonds, A.S., Gersch, J., Ammouri, H., et al. (2018). The ModERN Resource: Genome-
1211 Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription
1212 Factors. Genetics *208*, 937–949.

1213 Kuhn, R., Kuhn, C., Börsch, D., Glätzer, K.H., Schäfer, U., and Schäfer, M. (1991). A cluster of
1214 four genes selectively expressed in the male germ line of Drosophila melanogaster. Mech. Dev.
1215 *35*, 143–151.

1216 Law, N.C., Oatley, M.J., and Oatley, J.M. (2019). Developmental kinetics and transcriptome
1217 dynamics of stem cell specification in the spermatogenic lineage. Nat. Commun. *10*, 2787.

1218 Leatherman, J.L., and DiNardo, S. (2010). Germline self-renewal requires cyst stem cells and
1219 stat regulates niche adhesion in Drosophila testes. Nat. Cell Biol. *12*, 806–811.

1220 Li, H., Janssens, J., De Waegeneer, M., Sai, S.K., Davie, K., Gardeux, V., Saelens, W., Fabrice,
1221 D., Brbić, M., Spanier, K., et al. (2022). Fly Cell Atlas: A single-nucleus transcriptomic atlas of
1222 the adult fruit fly. Science *375*, eabk2432.

1223 Li, M.A., Alls, J.D., Avancini, R.M., Koo, K., and Godt, D. (2003). The large Maf factor Traffic
1224 Jam controls gonad morphogenesis in Drosophila. Nat. Cell Biol. *5*, 994–1000.

1225 Liu, Y., Li, J.S.S., Rodiger, J., Comjean, A., Attrill, H., Antonazzo, G., Brown, N.H., Hu, Y., and
1226 Perrimon, N. (2022). FlyPhoneDB: an integrated web-based resource for cell–cell
1227 communication prediction in Drosophila. Genetics *220*, iyab235.

1228 Lowe, N., Rees, J.S., Roote, J., Ryder, E., Armean, I.M., Johnson, G., Drummond, E., Spriggs,
1229 H., Drummond, J., Magbanua, J.P., et al. (2014). Analysis of the expression patterns,
1230 subcellular localisations and interaction partners of Drosophila proteins using a pigP protein trap
1231 library. Development *141*, 3994–4005.

1232 Lu, D., Sin, H.S., Lu, C., and Fuller, M.T. (2020). Developmental regulation of cell type-specific
1233 transcription by novel promoter-proximal sequence elements. Genes Dev. *34*, 663–677.

1234 Mahadevaraju, S., Fear, J.M., Akeju, M., Galletta, B.J., Pinheiro, M.M.L.S., Avelino, C.C.,
1235 Cabral-de-Mello, D.C., Conlon, K., Dell'Orso, S., Demere, Z., et al. (2021). Dynamic sex
1236 chromosome expression in Drosophila male germ cells. Nat. Commun. *12*, 892.

1237 McLaughlin, C.N., Brbić, M., Xie, Q., Li, T., Horns, F., Kolluru, S.S., Kebschull, J.M., Vacek, D.,
1238 Xie, A., Li, J., et al. (2021). Single-cell transcriptomes of developing and adult olfactory receptor
1239 neurons in Drosophila. Elife *10*, e63856.

1240 Michel, M., Kupinski, A.P., Raabe, I., and Bökel, C. (2012). Hh signalling is essential for somatic
1241 stem cell maintenance in the Drosophila testis niche. Development *139*, 2663–2669.

1242 Morris, C.A., Benson, E., and White-Cooper, H. (2009). Determination of gene expression
1243 patterns using in situ hybridization to Drosophila testes. Nat. Protoc. *4*, 1807–1819.

1244 Noguchi, T., Koizumi, M., and Hayashi, S. (2011). Sustained elongation of sperm tail promoted
1245 by local remodeling of giant mitochondria in Drosophila. Curr. Biol. *21*, 805–814.

1246 Okegbe, T.C., and DiNardo, S. (2011). The endoderm specifies the mesodermal niche for the
1247 germline in Drosophila via Delta-Notch signaling. Development *138*, 1259–1267.

1248 Olivieri, G., and Olivieri, A. (1965). Autoradiographic study of nucleic acid synthesis during
1249 spermatogenesis in Drosophila melanogaster. Mutat. Res. Mol. Mech. Mutagen. *2*, 366–380.

1250 Papagiannouli, F., and Mechler, B.M. (2009). discs large regulates somatic cyst cell survival and
1251 expansion in Drosophila testis. Cell Res. *19*, 1139–1149.

1252 Plass, M., Solana, J., Wolf, F.A., Ayoub, S., Misios, A., Glažar, P., Theis, F.J., Kocks, C., and
1253 Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell
1254 transcriptomics. *1723*, 1–17.

1255 Qadir, M.M.F., Álvarez-Cubela, S., Klein, D., van Dijk, J., Muñiz-Anquela, R., Moreno-
1256 Hernández, Y.B., Lanzoni, G., Sadiq, S., Navarro-Rubio, B., García, M.T., et al. (2020). Single-
1257 cell resolution analysis of the human pancreatic ductal progenitor cell niche. Proc. Natl. Acad.
1258 Sci. *117*, 10876–10887.

1259 Raja, J.S., and Renkawitz-Pohl, R. (2006). Replacement by Drosophila melanogaster
1260 Protamines and Mst77F of Histones during Chromatin Condensation in Late Spermatids and
1261 Role of Sesame in the Removal of These Proteins from the Male Pronucleus. Mol. Cell. Biol. *26*,
1262 3682.

1263 Rust, K., Byrnes, L.E., Yu, K.S., Park, J.S., Sneddon, J.B., Tward, A.D., and Nystul, T.G.
1264 (2020). A single-cell atlas and lineage analysis of the adult Drosophila ovary. Nat. Commun. *11*.

1265 Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction
1266 of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

1267 Schäfer, M., Kuhn, R., Bosse, F., and Schäfer, U. (1990). A conserved element in the leader

1268 mediates post-meiotic translation as well as cytoplasmic polyadenylation of a Drosophila
1269 spermatocyte mRNA. EMBO (Eur. Mol. Biol. Organ.) J. *9*, 4519–4525.

1270 Schäfer, M., Nayernia, K., Engel, W., and Schäfer, U. (1995). Translational Control in
1271 Spermatogenesis. Dev. Biol. *172*, 344–352.

1272 Schaum, N., Karkanias, J., Neff, N.F., May, A.P., Quake, S.R., Wyss-Coray, T., Darmanis, S.,
1273 Batson, J., Botvinnik, O., Chen, M.B., et al. (2018). Single-cell transcriptomics of 20 mouse
1274 organs creates a Tabula Muris. Nature *562*, 367–372.

1275 Sebé-Pedrós, A., Saudemont, B., Chomsky, E., Plessier, F., Mailhé, M.P., Renno, J., Loe-Mie,
1276 Y., Lifshitz, A., Mukamel, Z., Schmutz, S., et al. (2018). Cnidarian Cell Type Diversity and
1277 Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. Cell *173*, 1520-1534.e20.

1278 She, M., Tang, M., Jiang, T., and Zeng, Q. (2021). The Roles of the LIM Domain Proteins in
1279 Drosophila Cardiac and Hematopoietic Morphogenesis . Front. Cardiovasc. Med. *8*.

1280 Shi, Z., Lim, C., Tran, V., Cui, K., Zhao, K., and Chen, X. (2020). Single-cyst transcriptome
1281 analysis of Drosophila male germline stem cell lineage. Development *147*, dev184259.

1282 Shiraiwa, T., Nitasaka, E., and Yamazaki, T. (2000). Geko, a Novel Gene Involved in Olfaction
1283 in Drosophila Melanogaster. J. Neurogenet. *14*, 145–164.

1284 Shivdasani, A.A., and Ingham, P.W. (2003). Regulation of Stem Cell Maintenance and Transit
1285 Amplifying Cell Proliferation by TGF-β Signaling in Drosophila Spermatogenesis. Curr. Biol. *13*,
1286 2065–2072.

1287 Siebert, S., Farrell, J. a., Cazet, J.F., Abeykoon, Y., Primack, A.S., Schnitzler, C.E., and Juliano,
1288 C.E. (2019). Stem cell differentiation trajectories in Hydra resolved at single-cell resolution.
1289 Science *365*, 1–8.

1290 Sitnik, J.L., Francis, C., Hens, K., Huybrechts, R., Wolfner, M.F., and Callaerts, P. (2014).
1291 Neprilysins: An Evolutionarily Conserved Family of Metalloproteases That Play Important Roles
1292 in Reproduction in Drosophila. Genetics *196*, 781–797.

1293 Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki,
1294 M., Nef, S., Gnirke, A., et al. (2013). Cellular Source and Mechanisms of High Transcriptome
1295 Complexity in the Mammalian Testis. Cell Rep. *3*, 2179–2190.

1296 Susic-Jung, L., Hornbruch-Freitag, C., Kuckwa, J., Rexer, K.-H., Lammel, U., and Renkawitz-
1297 Pohl, R. (2012). Multinucleated smooth muscles and mononucleated as well as multinucleated
1298 striated muscles develop during establishment of the male reproductive organs of Drosophila
1299 melanogaster. Dev. Biol. *370*, 86–97.

1300 Tanentzapf, G., Devenport, D., Godt, D., and Brown, N.H. (2007). Integrin-dependent anchoring
1301 of a stem-cell niche. Nat. Cell Biol. *9*, 1413–1418.

1302 Tazuke, S.I., Schulz, C., Gilboa, L., Fogarty, M., Mahowald, A.P., Guichet, A., Ephrussi, A.,
1303 Wood, C.G., Lehmann, R., and Fuller, M.T. (2002). A germline-specific gap junction protein
1304 required for survival of differentiating early germ cells. Development *129*, 2529–2539.

1305 Thor, S., and Thomas, J.B. (1997). The Drosophila islet Gene Governs Axon Pathfinding and
1306 Neurotransmitter Identity. Neuron *18*, 397–409.

1307 Tokuyasu, K.T. (1975). Dynamics of spermiogenesis in Drosophila melanogaster. VI.
1308 Significance of “onion” nebenkern formation. J. Ultrastruct. Res. *53*, 93–112.

1309 Tokuyasu, K.T., Peacock, W.J., and Hardy, R.W. (1972). Dynamics of spermiogenesis in
1310 Drosophila melanogaster. II. Coiling process. Z. Zellforsch. Mikrosk. Anat. *127*, 492–525.

1311 Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak,
1312 K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions
1313 are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

1314 Tulina, N., and Matunis, E. (2001). Control of Stem Cell Self-Renewal in Drosophila
1315 Spermatogenesis by JAK-STAT Signaling. Science *294*, 2546–2549.

1316 Vedelek, V., Bodai, L., Grézal, G., Kovács, B., Boros, I.M., Laurinyecz, B., and Sinka, R. (2018).
1317 Analysis of Drosophila melanogaster testis transcriptome. BMC Genomics *19*, 697.

1318 Voog, J., Sandall, S.L., Hime, G.R., Resende, L.P.F., Loza-Coll, M., Aslanian, A., Yates, J.R.,
1319 Hunter, T., Fuller, M.T., and Jones, D.L. (2014). Escargot Restricts Niche Cell to Stem Cell
1320 Conversion in the Drosophila Testis. Cell Rep. *7*, 722–734.

1321 Wasbrough, E.R., Dorus, S., Hester, S., Howard-Murkin, J., Lilley, K., Wilkin, E., Polpitiya, A.,
1322 Petritis, K., and Karr, T.L. (2010). The Drosophila melanogaster sperm proteome-II (DmSP-II).
1323 J. Proteomics *73*, 2171–2185.

1324 Wendt, G., Zhao, L., Chen, R., Liu, C., O’Donoghue, A., Caffrey, C., Reese, M., and Collins, J.
1325 (2020). A single-cell RNA-seq atlas of Schistosoma mansoni identifies a key regulator of blood
1326 feeding. Science *369*, 1644–1649.

1327 White-Cooper, H., Schafer, M.A., Alphey, L.S., and Fuller, M.T. (1998). Transcriptional and
1328 post-transcriptional control mechanisms coordinate the onset of spermatid differentiation with
1329 meiosis I in Drosophila. Development *125*, 125–134.

1330 Wilk, R., Hu, J., and Krause, H.M. (2017). In Situ Hybridization: Fruit Fly Embryos and Tissues.
1331 Curr. Protoc. Essent. Lab. Tech. *15*, 9.3.1-9.3.26.

1332 Witt, E., Benjamin, S., Svetec, N., and Zhao, L. (2019). Testis single-cell RNA-seq reveals the
1333 dynamics of de novo gene transcription and germline mutational bias in drosophila. Elife *8*, 1–
1334 22.

1335 Zoller, R., and Schulz, C. (2012). The Drosophila cyst stem cell lineage. Spermatogenesis *2*,
1336 145–157.

1337

1338