Developmental Regulation of Alternative Polyadenylation in an Adult Stem Cell Lineage

Lorenzo Gallicchio[1], Neuza R. Matias[1], Fabian Morales-Polanco[2,3], Iliana Nava[1], Sarah Stern[1], Yi Zeng[3], Margaret T. Fuller[1,3,C]

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford USA.
[2]Department of Biology, Stanford University School of Humanities and Sciences, Stanford USA.
[3]Department of Genetics, Stanford University School of Medicine, USA.

[C]Correspondence to mtfuller@stanford.edu

Running Title: Developmental regulation of alternative polyadenylation

Keywords: Alternative Polyadenylation, Development, Cellular Differentiation, Spermatogenesis, mRNA processing, Cleavage factor Complex II.

Abstract

Co-transcriptional alternate processing of nascent mRNA molecules can make major contributions to cell type specific gene expression programs as proliferating precursor cells initiate terminal differentiation. Alternative Cleavage and Polyadenylation (APA) can result in the production of mRNA isoforms from the same gene locus with either longer or shorter 3'UTRs. In *Drosophila* spermatogenesis, approximately 500 genes undergo APA as proliferating spermatogonia differentiate into spermatocytes, producing transcript isoforms with shortened 3'UTRs, and resulting in profound stage specific changes in the proteins expressed. The molecular mechanisms that specify usage of upstream polyadenylation sites in spermatocytes are thus key to understanding the changes in cell state. Here, we show that PCF11 and Cbc, the two components of Cleavage factor II (CFII), orchestrate APA switching during *Drosophila* spermatogenesis. Knockdown of *PCF11* or *cbc* in spermatocytes caused dysregulation of APA, with many transcripts normally cleaved at a proximal site in spermatocytes now cleaved at their distal site, as in spermatogonia. Although *PCF11* is widely expressed, *cbc* is strongly upregulated in spermatocytes. Our findings reveal a developmental mechanism where changes in activity of specific cleavage factors can direct cell type specific APA at selected genes, presenting CFII as a key developmental regulator of APA during spermatogenesis.

Introduction

Alternative cleavage and polyadenylation (APA) of mRNAs results from changes in the site at which the 3' end cut that terminates nascent transcripts is made (Tian and Manley 2016; Gallicchio *et al.* 2023). If alternative polyadenylation sites are located in the 3'UTR of a genetic locus, APA results in production of alternate mRNA isoforms that differ in 3'UTR length and content, with potentially profound effects on transcript fate, including translation, stability, and localization (An *et al.* 2008; Mayr and Bartel 2009; Mansfield and Keene 2012; Berry *et al.* 2022). Since transcripts from more than 70% of human genes are alternatively polyadenylated under one condition or another, APA is widespread (Gruber and Zavolan 2019). The most intriguing cases are those where the choice of 3' end cut site is developmentally regulated, leading, for example, to production of transcripts with longer 3'UTRs from specific genes in neurons or shorter 3'UTRs in differentiating male germ cells compared to precursor cells (Hilgers *et al.* 2012; Miura *et al.* 2013; Shan *et al.* 2017; Vallejos Baier *et al.* 2017; Grassi *et al.* 2019; Berry *et al.* 2022). Changes in 3'UTR length due to APA have also been observed in many pathological states, such as the 3'UTR shortening of many transcripts observed in some cancers (Mayr and Bartel 2009; Morris *et al.* 2012; Venkat *et al.* 2020). Because these changes in 3'UTR content can have dramatic effects on protein expression and so cell state, cell type specific regulation of APA is emerging as an important developmental mechanism to promote robust cell fate decisions.

During cleavage and polyadenylation, the last step of mRNA synthesis, the nascent mRNA molecule is cleaved and a string of adenosines is added by nuclear PolyA polymerase to protect the newly made transcript from degradation and facilitate its export from the nucleus (Jing *et al.* 1999). Most mRNAs undergo cleavage and polyadenylation, with a notable exception being histone transcripts, which are not polyadenylated (Marzluff *et al.* 2008). Cleavage and polyadenylation is mediated by the cleavage machinery, which is composed of 4 well conserved multi-protein-complexes: the core complex CPSF (Cleavage and Polyadenylation Factor), which contains the endonuclease (CPSF73) that makes the actual cut that forms the transcript 3' end, CstF (Cleavage and Stimulation Factor), CFI (Cleavage Factor I) and CFII (Cleavage Factor II). The cleavage machinery is also associated with several non-complex proteins: Symplekin (a scaffolding protein), PABP2 and the nuclear PolyA polymerase (Figure 1 A) (Gallicchio *et al.* 2023).

76 The site at which 3' end cleavage will take place is guided by recognition by CPSF through the
77 components CPSF30 and WDR33 of the polyadenylation signal (PAS), an hexamer (most
78 commonly AAUAAA or a close variant) in the nascent mRNA usually ~15-30 nucleotides upstream
79 of the cleavage site (Clerici *et al.* 2018; Gallicchio *et al.* 2023). While the central role of the core
80 complex CPSF in recognizing the PAS and cleaving the nascent mRNA molecule is relatively clear,
81 the specific contributions of the other complexes and associated factors in specifying whether,
82 where, on what specific gene products, and in which cell types a cut is made is under investigation
83 (Vallejos Baier *et al.* 2017; Zhu *et al.* 2018; Turner *et al.* 2020).
84 Developmentally regulated changes in the PAS site utilized can have profound effects on
85 the suite of proteins expressed as cells progress from one state to the next in a developmental
86 sequence. For example, during spermatogenesis in flies and mice, many genes express mRNA
87 isoforms with long 3'UTRs in proliferating spermatogonia, but isoforms with shorter 3'UTRs at
88 later stages of germ cell development (Li *et al.,* 2016; Lee *et al.,* 2022; Berry *et al.,* 2022). In
89 *Drosophila melanogaster* testes (Figure 1 B), over 500 genes express transcripts cleaved at a distal
90 PAS in proliferating spermatogonia but processed at a more proximal PAS in differentiating
91 spermatocytes, resulting in production of mRNA isoforms with much shorter 3'UTRs (Figure 1 C).
92 Spermatogenesis in *Drosophila* starts with an asymmetric cell division that causes one stem cell
93 daughter to detach from hub cells, stop self-renewing and start mitotic proliferation, founding a
94 line of interconnected spermatogonia enclosed in two cysts cells. After 4 rounds of mitosis the
95 spermatogonia in the cyst start differentiation, beginning with pre meiotic S phase, and becoming
96 spermatocytes. After roughly 3.5 days of meiotic prophase, during which the germ cells grow 25
97 times in volume, the spermatocytes undergo meiosis I and II, producing a cyst of 64 haploid
98 spermatids. These then undertake the dramatic morphological changes that lead to production
99 of mature and functional sperm (Figure 1 B). The stage specific 3'UTR shortening due to APA
100 observed in spermatocytes compared to spermatogonia (Figure 1 C) resulted in a dramatic shift
101 in the proteins expressed, with transcripts from at least 200 genes switching from co-migrating
102 with polysomes in spermatogonia to not associated with ribosomes in young spermatocytes. In
103 addition, transcripts from another 50 genes were not translated in spermatogonia but associated
104 with polysomes in spermatocytes (Berry *et al.,* 2022). Thus, the developmental program(s) that

105 specify whether nascent transcripts are cleaved at proximal 3'UTR sites in spermatocytes rather
106 than the distal sites utilized in spermatogonia can switch expression of many proteins from on in
107 spermatogonia to off in spermatocytes or vice versa, with potential impacts on cell state.

108 Here we show that cell type specific production of transcripts with short 3'UTRs from
109 specific genes due to APA in spermatocytes is regulated by one of the 4 complexes involved in 3'
110 end cleavage and polyadenylation. We found that the two components of CFII, *PCF11* and
111 *crowded by cid (cbc)*, are up regulated in spermatocytes compared to spermatogonia. When
112 levels of either *PCF11* or *cbc* were knocked down (KD) in spermatocytes by cell type specific RNAi,
113 over 230 genes (41%) for which nascent transcripts would normally be cleaved at a proximal PAS
114 in spermatocytes resumed producing transcripts cleaved at the distal PAS, utilized in
115 spermatogonia. Lowering protein levels of components of the cleavage machinery associated
116 complexes CFI or CstF had much less effect on APA at these loci, suggesting that CFII plays a special
117 role in regulating APA during this differentiation sequence. The PCF11 and Cbc proteins interact
118 structurally: immunoprecipitation of Cbc from testis extracts brought with it a >250KD form of
119 PCF11. Reduced expression of *PCF11* protein in spermatocytes due to RNAi resulted in
120 concomitant loss of Cbc, suggesting that Cbc protein was not stable without its binding partner
121 in CFII. However, the reciprocal was not true. Plentiful PCF11 protein persisted in the nuclei of
122 spermatocytes in which expression of *cbc* had been knocked down by RNAi. Strikingly, *PCF11*
123 mRNA is broadly expressed in many cell types in adult flies, while *cbc* expression is low in most
124 adult tissues but up regulated in spermatocytes. These observations suggest a model where a
125 developmentally regulated increase in expression of *cbc* in spermatocytes leads to greater activity
126 of cleavage factor CFII. This in turn facilitates recognition of an upstream PAS, and consequent
127 3'UTR shortening by APA of nascent transcripts from specific genes, indicating CFII as a key player
128 in developmental regulation of alternative cleavage and polyadenylation.

Results

Specific cleavage factor proteins are upregulated in spermatocytes.

Immunostaining of whole mount *Drosophila* testes revealed that some but not all cleavage factor proteins are up-regulated in spermatocytes (Figure 1 D-J). Relative expression levels of six protein components of various sub-complexes of the cleavage factor machinery were measured by immunostaining using available antibodies against endogenous proteins (for CstF50, CstF64, PABP2 and PCF11), against HA (for a CRISPR *Drosophila* line tagged with HA at the endogenous *cbc* locus by CRISPR, *HA:cbc*), or against GFP (for a *Drosophila* line carrying a Fosmid based transgenic copy of CPSF6 tagged with sfGFP, (Sarov *et al.* 2016)). In comparing nuclear signal in spermatogonia and spermatocytes, each measurement was divided by the area of the corresponding nucleus to account for the great difference in nuclear size between the two cell types. Immunofluorescence signal from antibodies recognizing two members of the CstF complex, CstF50 (Figure 1 E) and CstF64 (Figure 1 F), decreased slightly or remained almost the same in nuclei in spermatocytes compared to spermatogonia (quantified in Figure 1 D). In contrast, immunofluorescence signal from antibodies recognizing PABP2 (Figure 1 G), the CFI component CPSF6 (Figure 1 H) and the two members of CFII, Cbc (Figure 1 I) and PCF11 (Figure 1 J) increased in nuclei from spermatocytes compared to spermatogonia (quantified in Figure 1 D).

Down-regulation of CFII components in spermatocytes shifted APA to more distal sites.

To knock down (KD) expression of cleavage factor components in spermatocytes, females carrying a *bamGal4* expression driver coupled with *UAS>dicer2* were crossed with males carrying RNAi constructs under control of a *UAS* array with an *hsp70* promoter. *bamGal4* flies express Gal4 under control of the *bag-of-marble (bam)* promoter, which fires in mid to late spermatogonial and early spermatocyte stages. The crosses were incubated at 25°C for three days, then the progeny shifted to 29°C to facilitate effective transcriptional activation by Gal4.

Immunofluorescence staining revealed that *bamGal4* driven KD using the *VDRC103710* RNAi line lowered levels of *PCF11* protein detected in spermatocyte nuclei to a level similar to that detected in spermatogonia (Figure 2 B). Testes from *UAS>PCF11-RNAi-VDRC-103710;bamGal4* males viewed in live squash preparations by phase contrast microscopy had

158 plentiful spermatocytes but few elongating spermatid bundles (Figure 2 E). An RNAi line directed
159 against a different region of the PCF11 transcript, *UAS*>PCF11-RNAi-*NIG-10228R-4*, also lowered
160 the level of anti-PCF11 immunofluorescence signal in spermatocytes when driven using *bamGal4*
161 (Figure 2 C). Consistent with the less complete KD, testes from *UAS*>PCF11-RNAi-*NIG-10228R-*
162 *4;bamGal4* males viewed by phase contrast microscopy had some elongating spermatid bundles
163 (Figure 2 F). In both cases, PCF11 protein was still detected in early spermatogonia and in nuclei
164 of the somatic cyst cells that accompany spermatocyte cysts, as expected since the *bamGal4*
165 driver is not expressed in these cells (Figure 2 B, C, arrows).

166 KD of *cbc*, the partner of PCF11 in CFII, also caused defects in spermatogenesis.
167 Immunofluorescence staining of testes from males heterozygous for a *cbc* allele tagged at the
168 endogenous locus with the HA epitope (*HA:cbc*) showed strong anti-HA signal in spermatocyte
169 nuclei, compared to little or no signal detected in spermatogonia. However, for two RNAi lines
170 directed against different regions of *cbc*, KD under control of *bamGal4* reduced the level of anti-
171 HA signal in spermatocytes to below the level of detection, as in spermatogonia (Figure 2 G-H-I).
172 In both cases, analysis of unfixed squash preparations by phase contrast microscopy showed that
173 the KD testes had abundant spermatocytes but lacked elongating spermatid bundles (Figure 2 K-
174 L).

175 Hybridization chain reaction fluorescence *in situ* hybridization (HCR-FISH) revealed that
176 KD of *PCF11* or *cbc* in spermatocytes by RNAi affected APA at both the *Kpc2* (also named
177 *isopeptidase-T-3*) and *Red* genes (Figure 3). *Kpc2* and *Red* had both been found to express
178 transcripts with a long 3'UTR in testis filled with spermatogonia but a much shorter 3'UTR in
179 testes filled with mid-stage differentiating spermatocytes (Figure 3 A, E, J). Comparing signal in
180 the same sample from two different sets HCR-FISH probes (SOM table 1), one recognizing the
181 common coding sequence and one recognizing the 3'UTR extension (Figure 3 A), revealed the
182 proportion of the total transcripts from a given gene (recognized by the common probe set) that
183 contained the 3' UTR extension (recognized by the 3'UTR extension probe). For both *Kpc2* and
184 *Red*, KD of either *PCF11* or *cbc* in spermatocytes under control of *bamGal4* dramatically increased
185 the level of signal from the 3'UTR extension detected in spermatocytes (Figure 3 C, D, H, I)
186 compared to testes from control flies raised under the same temperature regimen (Figure 3 B, G).

187 Quantification of the images confirmed that the ratio of fluorescence signal from probes
188 recognizing the long 3'UTR versus from probes against the coding sequence (CDS) for each gene
189 was significantly higher in the *PCF11* or *cbc* KD spermatocytes than in spermatocytes from control
190 testes (Figure 3 F, K).

191 Analysis of cleavage site usage by 3'seq from testes in which either *PCF11* or *cbc* had been
192 knocked down in spermatocytes by RNAi under control of *bamGal4* showed that the increase in
193 signal from the 3'UTR extension observed in KD spermatocytes was due to a shift from cleavage
194 at the proximal site normally utilized in spermatocytes to the more distal cut site normally utilized
195 in spermatogonia for both *Kpc2* and *Red* (Figure 4 A, B). The effect of KD of *cbc* was strongest,
196 while the effect of knock down of *PCF11* with the *UAS>PCF11-RNAi-NIG-10228R-4* construct was
197 weakest (Figure 3 A, B), consistent with the less efficient reduction in PCF11 protein levels
198 observed with that particular line (Figure 2 C).

199 3'seq data revealed that loss of function of *cbc* or *PCF11* affected APA at many loci. For
200 roughly half of the >500 genes previously identified by Berry *et al.* (2022) as undergoing 3'UTR
201 shortening by APA in spermatocytes compared to spermatogonia, KD of either *PCF11* or *cbc* in
202 spermatocytes resulted in increased usage of the distal cleavage site normally used in
203 spermatogonia (Figure 4 C). 3'seq data was analyzed via PolyA miner (Yalamanchili *et al.* 2020)
204 providing as input the list of pairs of proximal PAS (pPAS) and distal PAS (dPAS) from the >500
205 genes most changed in spermatogonia versus spermatocytes called by Berry *et al.* (2022). Orange
206 indicates that the frequency of the gene being cleaved at the distal PAS in a KD sample was higher
207 than in control testes, while blue indicates that the frequency of the gene being cleaved at the
208 proximal PAS in a KD sample was higher than in control testes. Noticeably, a negative PolyA index
209 does not indicate a shortening event but indicates a gene for which the proximal PAS was
210 overrepresented in a KD compared to a control sample. This could be caused either by different
211 expression of the gene compared to the control or by the different cell composition, as generally
212 KDs do not have as many elongating spermatids as a control testes. Again, KD of *cbc* in
213 spermatocytes had the strongest effect on the APA switch, affecting the highest number of genes.
214 Importantly, >94% of the genes where APA was affected (PolyA index > 0.5) by KD of *PCF11* also
215 showed a switch in APA with KD of *cbc,* consistent with these two CFII components acting

216 together in the choice of where to make the 3' end cut for transcripts from specific genes. Analysis
217 of RNAseq libraries from testes samples in which *cbc* or *PCF11* where knocked down under
218 *bamGal4* revealed that in general the genes that had a PolyA index either greater than 0.5 or
219 smaller than negative 0.5 were not expressed at dramatically different levels in KD versus control
220 testes (SOM Figure 3).

221 3'seq also revealed that lowering the levels of the nuclear protein PABP2 in
222 spermatocytes also frequently affected APA (SOM Figure 4). Immunofluorescence staining
223 showed effective KD of PABP2 protein to below detection in spermatocytes from flies carrying
224 *UAS>PABP2-RNAi-VDRC-106466;bamGal4*, although PABP2 protein was still abundant in nuclei of
225 somatic cyst cells, as expected (SOM Figure 4 B). For two different RNAi lines, testes from *UAS-*
226 *PABP2-RNAi;bamGal4* flies had plentiful spermatocytes but lacked elongating spermatid bundles
227 (SOM Figure 4 D-E). 3'seq revealed that in testes from *UAS>PABP2-RNAi-VDRC-106466;bamGal4*,
228 a substantial percentage of transcripts from the *Kpc2* and *Red* loci switched 3' end cleavage from
229 the proximal site normally utilized in spermatocytes to the more distal site normally utilized in
230 spermatogonia (Figure 4 A, B). Global analysis of the 3'seq data showed that for the >500 genes
231 identified previously as undergoing APA in spermatocytes compared to spermatogonia, for >86%
232 of the genes that switched toward the more distal PAS (PolyA index > 0.5) after KD of *PABP2*, APA
233 was similarly affected (PolyA index > 0.5) by KD of either *PCF11* or *cbc* (Figure 4 C, F).

234 In contrast, KD of any of the three components of the CstF complex (*CstF64*, *CstF50*, and
235 *su(f)*) or a component of CFI (*CPSF6*) in spermatocytes had much less effect on the observed APA
236 profile of mRNAs expressed in testes assessed by 3'seq (Figure 4 C). Immunofluorescence staining
237 with antibodies against CstF64 or CstF50 showed that levels of the proteins were substantially
238 reduced in spermatocytes in testes in which expression of the respective genes had been KD by
239 RNAi under control of *bamGal4* (SOM Figure 5). In both cases, signal remained in the nuclei of
240 early spermatogonia and cyst cells, as expected from the cell type specificity of the *bamGal4*
241 expression driver. No antibody reagents against *su(f)* were available to assess effectiveness of the
242 RNAi construct directed against *su(f)*. However, the effects of knocking down *su(f)* on protein
243 levels of CstF64 and CstF50 in spermatocytes nuclei (SOM Figure 6), and the defects in
244 spermatogenesis in *su(f)-RNAi;bamGal4* testes observed by phase contrast microscopy (SOM

245 Figure 5), suggested that the KD of *su(f)* was effective. Levels of *CstF64* protein were substantially
246 reduced and levels of *CstF50* protein were below the level of detection by immunofluorescence
247 staining in spermatocytes from males carrying an RNAi construct directed against *su(f)* expressed
248 under control of *bamGal4* (SOM Figure 6 C, D). KD of *CstF64* under control of *bamGal4* also
249 lowered the levels of CstF50 in spermatocyte nuclei to below detection (SOM Figure 6 F). Notably,
250 when either *su(f)* or *CstF64* were KD in spermatocytes, nuclear CstF50 staining disappeared,
251 although some signal remained in the cytoplasm. Together the protein expression data suggest
252 that components of the CstF complex may depend on each other for protein stability and/or
253 nuclear localization, as previously reported (Grozdanov *et al.* 2018). Consistent with the KDs
254 affecting function, squash preparations of testes from *UAS>su(f)-RNAI-BL-65693;bamGal4*
255 showed abundant spermatocytes but no elongating spermatid stages (SOM Figure 5 G). Testes
256 from *UAS>CstF64-RNAi-BL65987; bamGal4* males also showed defective spermatogenesis, with
257 plentiful spermatocytes but extremely aberrant elongation stages (SOM Figure 5 F). Surprisingly,
258 even though CstF50 protein was clearly strongly reduced in spermatocytes from *UAS>CstF50-*
259 *RNAi-BL77377;bamGal4* males, spermatogenesis appeared largely normal by phase contrast
260 microscopy (SOM Figure 5 C).

261 KD of either *CstF64* or *su(f)* in spermatocytes shifted the site of 3' end formation toward
262 the more distal site utilized in spermatogonia at a very small percentage of the >500 genes
263 previously identified as subject to 3'UTR shortening in spermatocytes due to developmentally
264 regulated APA (Figure 4 C). These tended to be among the same set of genes where APA was
265 affected by KD of *cbc* or *Pcf11* (components of CFII): >88% of the genes affected by KD of *CstF64*
266 were also affected by KD of either *PCF11, cbc,* or both (Figure 4 G). However, the effects of loss of
267 function of CstF components on APA in spermatocytes was much less than the effect of loss of
268 CFII components: only 14 of the 235 genes where APA was strongly affected by KD of either *cbc*
269 or *PCF11* in spermatocytes also showed a switch toward the more distal cut site in testes where
270 function of *CstF64* had been knocked down in spermatocytes (Figure 4 G). All the genes affected
271 (PolyA index > 0.5) by *CstF64* KD were also affected either by *CstF50* or *su(f)* KDs. Interestingly, of
272 the few genes affected by *CstF50* KD, for almost all the choice of 3' end cut site was not affected
273 by *CstF64* or *su(f)* KDs (Figure 4 H). KD under control of *bamGal4* of *CPSF6*, a component of the

274 CFI complex, which binds nascent transcripts upstream the PAS, caused extensive defects in
275 spermatid differentiation for the two RNAi lines tested, suggesting loss of function (SOM Figure
276 7). Analysis of cut sites utilized by 3' end seq indicated that for the >500 genes previously
277 identified as undergoing APA in spermatocytes, the effect of *CPSF6* KD on APA was mild compared
278 to KDs of *cbc* or *PCF11,* although *CPSF6* KD affected APA at many of the same genes affected by
279 loss of function of CstF components (Figure 4 C).

280 *De novo* genome wide calling of PAS sites located on annotated 3'UTRs revealed that in
281 addition to many of the >500 genes from Berry *et al.* (2022), KD of components of the cleavage
282 machinery affected APA at additional loci (SOM Figure 8-9). For this analysis we only considered
283 genes in which the most up regulated PAS and the most down regulated PAS in a KD compared
284 to a WT sample were located in the annotated gene 3'UTR. Again, the strongest effects were
285 detected in *cbc* KD testes, followed by *PCF11* KD and *PABP2* KD testes (SOM Figure 8), while
286 *CstF64*, *CstF50*, *su(f)* and *CPSF6* KDs had a much milder effect on the global APA profile (SOM
287 Figure 9). When *cbc* or *PCF11* or *PABP2* were knocked down in spermatocytes many more genes
288 showed lengthening of 3'UTR, suggesting that expression levels of components of CFII and/or
289 PABP2 in spermatocytes helped functional cleavage at the proximal rather than the distal site.

290 Of the components of the cleavage machinery tested, the 3'seq data indicated that the
291 3'UTR shortening of transcripts from selected genes due to APA in spermatocytes was most
292 strongly dependent on levels of CFII (composed of *cbc* and *PCF11*), or *PABP2.* For almost half of
293 the >500 genes previously identified as undergoing the developmentally regulated APA, lowering
294 function of one or the other of these proteins in spermatocytes resulted in a shift toward usage
295 of the more distal 3' end cut site typical of spermatogonia. A number of genes seemed to be more
296 affected by KD of *cbc* than by KD of *PCF11*, for example *oaf* (Figure 4 D). However, over half of
297 the genes previously identified as subject to 3'UTR shortening in spermatocytes compared to
298 spermatogonia due to APA were not substantially affected by KD of any of the cleavage factor
299 machinery we tested, *nudE* for example (Figure 4 E and SOM Figure 2).
300
301
302

303 **Transcription of the CFII component *cbc* is strongly upregulated in spermatocytes.**

304 According to Fly Base, the *PCF11* locus can give rise to 7 different RNA isoforms (Figure 5
305 A), grouped in 3 classes for simplicity: a short isoform (*PCF11-RC*) of 819 nucleotides predicted to
306 encode a small protein (~10 KDa), a medium isoform (*PCF11-RB*) of 2647 nucleotides encoding a
307 ~63 KDa protein, and a group (*PCF11-RD-RE-RF-RG-RH*) of ~6000 nucleotide transcripts predicted
308 to encode a family of >200 KDa proteins that have subtle differences in a central region. All the
309 transcript isoforms share the same 5'UTR but differ in their 3'UTR (as well as in exon
310 combinations). PCF11 has a CID domain (POLII C-terminal binding domain) that is present in both
311 the PCF11-long and the 63 KDa PCF11-RB isoforms, but only partially present in the short protein
312 encoded by *PCF11-RC*. In addition, PCF11 have been shown to bind Cbc in mammalian and yeast
313 cells, through a region close to the C-terminus of PCF11 (Sadowski *et al.* 2003; Schäfer *et al.* 2018).
314 Notably, while the *PCF11-RNAi-NIG10228-R4* line targets a region common to all *PCF11* mRNA
315 isoforms, *PCF11-RNAi-VDRC107310* targets a region present only in the *PCF11-long* isoforms. KD
316 with *PCF11-RNAi-VDRC107310;bamGal4* in spermatocytes had a mildly stronger effect on the APA
317 profile detected by 3'seq than the KD with *PCF11-RNAi-NIG10228R-4;bamGal4*. The *cbc* locus is
318 simpler, with only one predicted transcript 1654 nucleotides long, encoding a 46.8 KDa predicted
319 protein (Figure 5 A).

320 The two components of CFII show substantially different transcription patterns in adult
321 flies. In the Fly Cell Atlas single nuclear RNA-seq analysis of tissues from adult flies (Li *et al.,* 2022),
322 transcription of *PCF11* was detected at generally high levels in many cell types (Figure 5 B). In
323 contrast, *cbc* expression was detected at substantial levels in only a small subset of nuclei from
324 adult tissues, with the clusters showing high *cbc* transcription belonging to the testes sample
325 (Figure 5 C, D). Plotting expression on the UMAP of snRNA-Seq data from testis plus seminal
326 vesicle revealed that *PCF11* is transcribed in many somatic cell types in the testis, including
327 terminal epithelia, seminal vesicle, nuclei from testis muscle cells and prominently in the somatic
328 cyst cells that enclose germ line cysts. *PCF11* transcripts were also expressed in germ cells, with
329 the level increasing in spermatocytes compared to spermatogonia (Figure 5 E). In contrast,
330 although transcription of *cbc* increased substantially in spermatocytes compared to
331 spermatogonia, little signal from *cbc* was detected in somatic cells in the testis.

332 HCR-FISH on testes revealed that long form transcript(s) from *PCF11*, were up-regulated
333 in spermatocytes compared to spermatogonia (Fig. 5 H-H'-H''). In contrast, the intermediate sized
334 transcript isoform *PCF11-RB*, detected with a set of probes recognizing the 3'UTR of *PCF11-RB*,
335 which derives from an intronic region of the *PCF11-long* isoforms so should not recognize fully
336 processed *PCF11-long* isoforms in the cytoplasm, was extremely low in both cell types (Fig. 5 I-I'-
337 I''). Single nuclear RNAseq from the Fly cell Atlas confirmed that while the collective *PCF11-long*
338 isoforms are highly expressed in the germline and in other cell lineages in the testis cluster
339 (including somatic cells and muscle cells, Figure 5 K) the *PCF11-RB* transcript is very lowly
340 expressed in the germline, and modestly expressed in the somatic and muscle lineages (Figure 5
341 L). Similar HCR-FISH using a set of probes against the *cbc* protein coding region confirmed that
342 expression of *cbc* transcripts was strongly up regulated in spermatocytes compared to
343 spermatogonia (Fig. 5 J-J'-J'').
344
345 **PCF11 and Cbc co-immunoprecipitate and knock down of *PCF11* decreases Cbc protein in**
346 **spermatocytes nuclei.**
347 Analysis of proteins expressed in wild type testes by polyacrylamide gel electrophoresis
348 (PAGE) followed by Western blot probed with antibodies against PCF11 showed a predominant
349 high molecular weight (>250 KDa) band (Figure 6 A), consistent with the expression of *PCF11-long*
350 mRNA isoforms detected by HCR-FISH and snRNA-seq (Figure 5 H, K). Anti-PCF11 also detected a
351 ~70 KDa protein isoform in testis extracts, as predicted from the *PCF11-RB* mRNA isoform. The
352 antibody used was raised against amino acids 1-283 of *Drosophila* PCF11 (Zhang and Gilmour
353 2006), however no protein product corresponding to PCF11-RC was observed, despite a
354 seemingly high level of the *PCF11-RC* transcript isoform detected by snRNAseq of *Drosophila*
355 testis tissue (SOM Figure 10). Western Blots of testes from homozygous flies carrying the *HA:cbc*
356 allele probed with anti-HA showed expression of the expected ~50 KDa isoform (Figure 6 B).
357 Immunoprecipitation with anti-HA antibodies from extracts of testes from flies
358 homozygous for *HA:cbc* co-immunoprecipitated the >250 KDa band from the PCF11 isoforms,
359 consistent with the Cbc and PCF11 proteins interacting in a complex (Figure 6 C). The 70 KDa
360 isoform of PCF11 was not detected in the immunoprecipitation, consistent with the prediction

361 that the long but not the medium or short PCF11 isoforms contain the Cbc binding site. Consistent
362 with interaction between Cbc and PCF11, accumulation of Cbc protein in spermatocytes required
363 PCF11. Analysis by RNA-seq showed that when PCF11 expression was knocked down in late
364 spermatogonia and spermatocytes by RNAi under control of *bamGal4*, *cbc* transcript was still
365 expressed in *PCF11* KD testes at a similar level compared to control testes (Figure 6 D). Likewise,
366 in testes undergoing KD of *cbc* by RNAi under control of *bamGal4*, transcripts from the *PCF11*
367 locus were still expressed (Figure 6 E). However, the situation was different at the level of protein
368 accumulation in spermatocytes. Simultaneous immunostaining of testis from *HA:cbc/+;bamGal4*
369 males with anti-PCF11 and anti-HA antibodies showed both proteins present in spermatocyte
370 nuclei from control males (Figure 6 F). However, for two different RNAi lines directed against
371 *PCF11*, KD of *PCF11* under control of *bamGal4* resulted in loss of the signal from HA:Cbc (Figure
372 6 G'-H', quantified in K). In contrast, KD of *HA:*cbc by RNAi under control of *bamGal4* did not cause
373 loss of signal from anti-PCF11 (Figure 6 I''-J'', quantified in L).

374 **Discussion**

375 Our results indicate that cell type specific differential expression of components of the
376 Cleavage Factor complex II (CFII) is a key developmental mechanism regulating alternative
377 polyadenylation as cells switch from proliferation to onset of differentiation in an adult stem cell
378 lineage. Cell type specific knock down (KD) experiments suggest that the developmentally
379 specified increase in expression of PCF11 and Cbc, the two components of CFII, is important for
380 the shift to 3' end formation at a more proximal cleavage site in spermatocytes compared to
381 spermatogonia at over 200 loci during *Drosophila* spermatogenesis. Knock down of several other
382 cleavage factors had much milder effects on shifting 3' end cleavage from the proximal site
383 normally utilized in spermatocytes to the more distal site normally utilized in spermatogonia,
384 singling out CFII for a key role in the developmentally regulated 3'UTR shortening due to APA in
385 normal *Drosophila* spermatocytes.

386 Of the two components of CFII, reducing levels of Cbc expressed in spermatocytes had the
387 most dramatic effect on PAS choice, effecting more than 40% of the >500 genes previously
388 identified as subject to APA leading to 3' end cleavage at a more proximal PAS in spermatocytes.
389 Expression of both *PCF11* and *cbc* is developmentally regulated, with mRNA and protein levels
390 considerably higher in spermatocytes compared to spermatogonia. Analysis of the single nuclear
391 RNAseq data from the Fly Cell Atlas project (Li *et al.,* 2022) revealed that expression of *cbc* is much
392 more cell type specific, being upregulated strongly in spermatocytes compared to other cell types
393 and tissues. In contrast, *PCF11* was expressed much more broadly, with transcription detected in
394 every fly tissue analyzed. This, plus our finding that knock down of *PCF11* in spermatocytes
395 resulted in failure of Cbc protein to accumulate, while PCF11 protein was still present in
396 spermatocytes where *cbc* was knocked down, suggest a model where PCF11 carries out functions
397 independent of Cbc, but when Cbc protein is upregulated in spermatocytes it binds some of the
398 PCF11 and recruits it into CFII, where it can bind and activate the core 3' end cleavage machinery
399 to cut at a more proximal PAS site in specific transcripts that are normally skipped in
400 spermatogonia. Although previous studies involving knock down of CLP1 (the mammalian
401 homolog of Cbc) in mammalian cells did not show remarkable effect on APA (Li *et al.* 2015;
402 Ogorodnikov *et al.* 2018), our findings raise the possibility that activity of Cbc in the choice of 3'

403 end cut site may be more important in some cell types than in others, and on certain specific
404 transcripts rather than in general.

405 Consistent with the model that PCF11 may carry out functions independent of CFII, PCF11
406 has been found to bind the CTD of RNA POLII *in vitro* and in yeast (Sadowski *et al.* 2003; Zhang
407 and Gilmour 2006), where it acts as a termination factor (Guéguéniat *et al.* 2017; Schäfer *et al.*
408 2018; Kamieniarz-Gdula *et al.* 2019). CHIP-seq of POLII in control versus *PCF11* knock down
409 human cells showed that POLII fails to terminate efficiently when PCF11 is reduced (Kamieniarz-
410 Gdula *et al.* 2019). Accordingly, downregulating *PCF11* expression levels has been shown to
411 induce 3'UTR lengthening in cancer cells and zebrafish (Li *et al.* 2015; Ogorodnikov *et al.* 2018;
412 Kamieniarz-Gdula *et al.* 2019). However, if/how skipping cleavage sites might be linked with POLII
413 termination, and what other functions PCF11 might carry out in association with RNA POLII still
414 need to be elucidated.

415 Cbc is a much smaller protein than PCF11, does not have a recognized RNA binding
416 domain, and immunoprecipitation of cbc *in vitro* failed to detect bound RNA (Schäfer *et al.* 2018).
417 On the other hand, PCF11 has been shown to bind RNA *in vitro*, with a preference for G-rich
418 regions. Our findings raise the possibility that Cbc might act as a developmentally regulated factor
419 that, when upregulated in certain cell types might bind to PCF11 to form CFII, utilize the ability of
420 PCF11 to recognize RNA, and modulate PCF11 activity to enhance 3' end cleavage at specific sites.

421 Developmentally regulated switches in the site at which nascent mRNAs are terminated,
422 such as those we have observed dependent on components of CFII, may serve as a mechanism
423 to quicky change the menu or levels of proteins expressed as cells progress through sequential
424 steps in differentiation. By effecting the length and content of 3'UTRs, APA can change whether
425 mRNAs are translated, blocked from translation, differentially localized, stabilized, or degraded.
426 Since APA is a co-transcriptional process that does not require opening or closing of chromatin or
427 re-initiation of transcription, it may provide a more rapid way to turn expression of some proteins
428 on and other proteins off, so changing cell state and canalizing cell fate decisions. As such,
429 understanding how the alternative choices in 3' end cut site are regulated is key for understanding
430 developmental regulation of cell states.

431 **Materials and Methods**

432

433 **Fly Strains and husbandry.**

434 Flies were grown on standard molasses media and kept at 25 °C, 29 °C or room temperature as
435 indicated. For RNAi induced knockdowns (KD), fly crosses (females *Bamgal4* driver X males RNAi
436 line) were kept at 25 °C for three days. Adults were then shifted to a new media bottle and kept
437 at 25 °C while eggs and early larvae were shifted to 29 °C to increase KD efficiency. To further
438 increase KD penetrance, we used a Gal4 driver that also expressed *UAS > dicer2*. Fly lines used in
439 this study: *W1118* (BL #3605), *Bam Gal4* (Chen and McKearin 2003), *CPSF6:GFP* (VDRC #318105),
440 *HA:Cbc* (N-terminal CRISPR tag, kindly gifted from the Wang lab, (Wu *et al.* 2021)), *PCF11*-RNAi_1
441 (VDRC #103710) *PCF11*_RNAi_2 (NIG #10228R4), *cbc*-RNAi_1 (BL #77426), *cbc*-RNAi_2 (VDRC
442 #100686), *CPSF6*-RNAi_1 (BL #34804), *CPSF6*-RNAi_2 (VDRC #107147), *CstF64*-RNAi_1 (BL
443 #65987), *CstF50*-RNAi_1 (BL #77377), *PABP2*-RNAi_1 (VDRC #106466), *PABP2*-RNAi_2 (VDRC
444 #33499), *su(f)*-RNAi_1 (BL 55693).

445

446 **Immunofluorescence (IF) staining and signal quantification.**

447 Testes from 0-3 days old were removed by dissecting in 1X PBS on a cyclops dissecting dish at
448 room temperature over 30 minutes or less, then testes were immediately fixed in an Eppendorf
449 with 5% formaldehyde for 30 minutes, with rocking at room temperature. If needed, fixed
450 samples were stored in 1X PBS at 4 °C for up to one week. Testes were then permeabilized in 0.3%
451 Triton-X-100 and 0.6% sodium deoxycholate for 30 minutes at room temperature and rinsed twice
452 in 1X PBS + 0.05% Tween-20 (PBT), blocked for 30 minutes in 1X Western Blocking Reagent (WBR,
453 Roche #) in PBT at room temperature, then incubated with primary antibody mixed in 1X
454 WBR+PBT overnight at 4 °C. After primary antibody incubation, testes were rinsed twice in PBT,
455 blocked again for 30 minutes in 1X WBR+PBT at room temperature, then incubated with
456 secondary antibody mixed in 1X WBR+PBT for 2 hours at room temperature in a dark container.
457 Testes were then rinsed twice in PBT and mounted on a glass slide in 10 uL of ProLong diamond
458 anti-fade with DAPI (ThermoFisher, CAT #P36962). Primary antibodies used: guinea pig anti-
459 CstF50 (1:300, kindly gifted from John T. Lis (Ni *et al.* 2008)), rabbit anti-PABP2 (1:500, kindly gifted

460 from Martine Simonelig (Benoit *et al.* 1999)), goat anti-Vasa (1:200, Santa Cruz Biotechnology sc-
461 26877, discontinued), rat anti-VASA (1:10 for supernatant or 1:100 for the concentrated version,
462 both from DSHB, antibody ID 760351) rabbit anti-CstF64 (1:400, kindly gifted from Zbigniew
463 Dominski (Skrajna *et al.* 2018)), rabbit anti-PCF11 (1:200, kindly gifted from David Scott Gilmour
464 (Zhang and Gilmour 2006)), rabbit anti-HA (1:500, Cell Signalling Technologies, #3724), mouse
465 anti-HA (1:500, CAT???), chicken anti-GFP (1:100, cat???). Secondary antibodies used: donkey or
466 goat anti-rabbit, anti-mouse, anti-guineapig, anti-chicken, anti-rat and anti-goat, all used at 1:500
467 and conjugated with either Alexa 488, Alexa 568, Alexa 594 or Alexa 647 depending on the
468 combination of antibodies used in a given experiment (all from the ThermoFisher catalog).
469 Imaging of IF experiments was performed on a Leica SP8 Confocal microscope. Protein levels were
470 quantified in Fiji by manually delineating the cell nucleui and applying the formula:
471 *Integrated_Density – (Area X Background_Mean)*.
472
473 **HCR^{TM}-RNA-FISH.**
474 Testes were dissected from 0-3 days old male flies in 1X PBS on a cyclops dissecting dish at room
475 temperature over a maximum of 30 minutes, then fixed and permeabilized as done for
476 immunofluorescence. The protocol for "generic sample in solution" from the Molecular
477 Instruments website was followed, hybridization buffer, wash buffer, amplification buffer and
478 hairpins are all from Molecular Instruments (HCR™ RNA-FISH Bundle). After permeabilization,
479 samples were incubated with 200 uL of probe hybridization buffer for 30 minutes at 37 ^{o}C. Probe
480 mix were prepared mixing 3.2 uL of 1 uM probes (one per gene or transcript) in 200 uL of probe
481 hybridization buffer. Hybridization buffer was removed from samples and probe mix added.
482 Samples were incubated overnight in probe mix at 37 ^{o}C. The following day samples were washed
483 4 times for 15 minutes each in 200 uL of probe wash buffer at 37 ^{o}C and then washed 3 times 5
484 minutes each in 5X SSCT (5X sodium chloride sodium citrate (SSC) and 0.1% Tween-20 in H_2O) at
485 room temperature. Samples were incubated in 500 uL of amplification buffer for 30 minutes at
486 room temperature. In the meantime hairpin mix was prepared: 30 pmol per sample (5 uL) of
487 harpin h1 and hairpin h2 (each fluorophore has its own harpin combination, in our case since we
488 always imaged 2 transcripts at the same time, we used 2 fluorophores and therefore 4 hairpins)

489 were heated at 95 °C for 90 seconds in a thermocycler in the dark and cool down at room
490 temperature in a drawer for 30 minutes (each hairpin in heat up and cooled in a separate tube).
491 Hairpin mix was prepared mixing 5 uL of each harpin to 250 uL of amplification buffer, and
492 samples were incubated in harpin mix at room temperature in the dark overnight. The following
493 day samples were washed with 500 uL 5X SSCT 2 times for 5 minutes each, then 2 times 30
494 minutes each and finally 1 time 5 minutes, always in the dark and at room temperature. Samples
495 were then mounted on a glass slide in 10 uL of ProLong diamond anti-fade with DAPI
496 (ThermoFisher, CAT #P36962). Probes were designed as in Bedbrook *et al.* 2023 and ordered from
497 Integrated DNA technologies as Oligo Pools (oPools™ 50 pmol/oligo). Harpins H1 and H2 were
498 ordered conjugated with either 488 or 546 fluorophores from Molecular Instruments. Probes
499 sequences are listed in supplementary material. Imaging was performed on a Leica SP8 confocal
500 microscope. For each testis, 5 areas from spermatocytes cytoplasm were quantified on Fiji from
501 both probe channels, subtracted by the background and then the ratio was calculated. The
502 following formula was used: *(Area_Mean_p2 – Back_Mean_p2) / (Area_Mean_p1 –*
503 *Back_Mean_p1)* where p2 is the probes set recognizing 3'UTR extension while p1 is the probe set
504 recognizing the gene CDS. Area's size varied between pictures but was constant between the 5
505 measurements from the same image.
506
507 **3' end sequencing and analysis.**
508 Testes were dissected from 0-3 days old male flies in S2 media in batches of 50-80 flies on a
509 cyclops dissecting dish at room temperature over a maximum of 1 hour per batch. Testes were
510 then immediately transfer in an Eppendorf containing 1X PBS, 1X PBS was removed and the testes
511 were snap-frozen in liquid nitrogen and stored at -80 °C. RNA was extracted from ~150 testes pairs
512 using the RNAeasy Mini kit (Qiagen, 74104) following the manufacturer directions with minor
513 adjustments. Frozen tissue was dissociated using a 1mL syringe with a 27G needle in RNAeasy
514 Mini kit lysis buffer supplied with 1:100 β-mercaptoethanol as suggested by the kit manual.
515 Quality of extracted RNA was assessed via Bioanalyzer, and the concentration measured by
516 Nanodrop and Qubit RNA HS Assay Kit (Life Technologies, Q32852). 3' end RNAseq libraries were
517 prepared using the Quant Seq (FWD) kit from Lexogen using 500 ng of total extracted RNA per

518 replicate. DNA libraries quality was checked via bioanalyzer and concentration measured with
519 Qubit dsDNA HS Assay Kit (Life Technologies, Q32854). Sequencing was performed by the
520 Stanford Genomics facility on an Illumina NextSeq 500 platform as in Berry *et al.* 2022. Each
521 biological sample was performed in duplicates. Raw reads were trimmed to remove adaptor
522 sequences and aligned to the *Drosophila* dm6 genome using bowtie 0.12.8 with a value of e=5000
523 to allow for the expected mismatched in the PolyA tail sequence, as in Berry *et al.* 2022.
524 Quantification of PAS usage of the ~500 genes identified in Berry *et al.* 2022 was performed using
525 PolyA miner (Yalamanchili *et al.* 2020), using as input a bed file containing the proximal and distal
526 PAS of the ~500 that are cleaved at a proximal PAS in spermatocytes compared to spermatogonia.
527 We generated this list by merging the proximal and distal cleavage sites from 3 different lists: bam^{-}
528 $^{/-}$ vs. 32H (very early spermatocytes), bam$^{-/-}$ vs. 48H (polar spermatocytes) and bam$^{-/-}$ vs 72H (mid-
529 stage maturing spermatocytes) to consider various stages of spermatocytes (32H, 48H and 72H
530 indicate hours after heat shock, which causes a pulse of bam expression in bam$^{-/-}$ flies, inducing
531 differentiation, as explained in Kim *et al.* (2017). As in this lists the coordinates of the cleavage
532 sites did not match perfectly, some genes had more than 2 cleavage sites considered. However,
533 PolyA miner was built to consider genes with multiple polyA sites.
534
535 ***De novo* PAS calling and APA analysis**.
536 For analysis of genome wide PAS usage changes between WT and KD samples, we adapted the
537 analysis pipeline from Zeng *et al.* 2024. The sequenced 3' end-seq libraries were adapter-
538 trimmed, and quality filtered using bbduk. The filtered reads were mapped to dm6 *Drosophila*
539 genome using STAR and then extracted for unique-mapping reads using Samtools. If a read was
540 mapped to a region that is immediately upstream of six consecutive As or of a 10-bp region with
541 at least 60% of As, it was considered as a mis-primed read and removed. The resulting filtered
542 reads were used for *de novo* PAS calling using LAPA[61] that used the default setting, except
543 requiring a replication rate cutoff at 0.95. For each gene, lengthening or shortening was quantified
544 by considering the ratio between the 2 highest changing PAS: the PAS that is more up-regulated
545 in the KD and the PAS that is more down-regulated in the KD. For this analysis we considered only
546 genes for which both PASs located in the 3'UTR.

547 **RNAseq and analysis.**

548 The same total RNA extracted for 3'seq was also used for preparation of short read RNAseq library.
549 1 ug of total extracted RNA was used to prepare libraries following the protocol from the NEBNext
550 Ultra II Direction RNA Library Prep Kit for Illumina. Each biological sample was performed in
551 duplicate. Quality and concentration of libraries were measured via Bioanalyzer and Qubit, as for
552 3'seq libraries. Sequencing was performed by Novagene on a NOVAseq, PE 150 Illumina Platform.
553 Raw reads were trimmed to remove the Illumina adapters using trim_galore, then aligned to the
554 *Drosophila* dm6 genome using STAR. Differential expression analysis was performed using the
555 package DEseq2.

556

557 **Fly Cell Atlas single nuclear RNAseq analysis**

558 Using the Fly Cell Atlas "Testis" polydT-primed 10x single-nuclear RNA sequencing dataset (Li *et*
559 *al.* 2023), we aggregated all reads from nuclei in each of the 111 Leiden 6.0 clusters (as in Berry
560 *et al.* 2022). For each cluster, reads that uniquely mapped to one of the isoform specific *PCF11*
561 3'UTRs were counted and normalized by the length of that 3'UTR as well as the total number of
562 reads derived from that cluster. This produced an isoform-specific rpkm value for each Leiden 6.0
563 cluster that was then log-normalized and plotted using Seurat. gene expression and cell type
564 annotation plots across all the adult *Drosophila* tissues analyzed in the Fly Cell Atlas project were
565 generated in Seurat using published code modified to visualize relevant genes and cell identities
566 of interest (Raz *et al.* 2023).

567

568 **Western Blot (WB).**

569 Testes from 20-50 0-3 days old male flies were quickly dissected into 1X PBS on a cyclops
570 dissecting dish at room temperature, transferred to an Eppendorf and snap froze in liquid
571 nitrogen and stored at -80 ^{o}C until needed. Frozen tissue was lysed in 40 uL of cold Lysis Buffer
572 (either RIPA buffer [50 mM TRIS pH 7.5, 250 mM NaCl, 0.1% SDS, 0.5% Sodium deoxycholate to
573 which was added one tablet per 50 uL of protease inhibitor cocktail (cOmplete ULTRA Tablets,
574 Mini, EDTA-free, EASY pack, Roche, REF 05 892 791 001)] or JUME Buffer [10 mM MOPS, 10 mM
575 EDTA, 8M UREA, 1% SDS]). Tissue was dissociated using a 1 mL syringe with a 27G needle, then

576 incubated rocking at 4 °C for 20 minutes. Following incubation, lysate was centrifuged at max
577 speed for 5 minutes at 4 degrees. Supernatant was transfered to a new tube, mixed with 4X Laemli
578 Buffer with β-mercaptoethanol, boiled for 7 minutes at 85 °C, then run on a Mini-PROTEAN TGX
579 Precast Protein Gel (4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 10-well, 50 µl #4561084).
580 The separated proteins were transferred from the gel to a PVDF membrane (BIO-RAD, Cat.#
581 1620177) overnight at 4 °C. The following day, the membrane was blocked overnight at 4 °C in 5%
582 powder nonfat milk dissolved in PBST (1X PBS with 0.1% Tween-20). After blocking, the
583 membrane was incubated with primary antibody mix in 2% milk for 1 hour at room temperature,
584 washed 3 times 20 minutes each with PBST at room temperature, incubated with secondary
585 antibody mix in 2% powder nonfat milk for one hour at room temperature and washed again 3
586 times 20 minutes each at room temperature. Stainings were then revealed using Revvity Health
587 Sciences Inc Western Lightning Plus-ECL, Enhanced Chemiluminescence Substrate (PerkinElmer).
588 Primary antibodies used for Western blots: Rabbit anti-HA (1:5000 or 1:1000, Cell Signalling
589 Technologies, #3724), Rabbit anti-PCF11 (1:10.000, kindly gifted from David Scott Gilmour (Zhang
590 and Gilmour 2006)). Secondary antibodies used for Western blots: Peroxidase AffiniPure™ Goat
591 Anti-Rabbit IgG (H+L) (1;10.000, Jackson ImmunoResearch AB_2313567).
592
593 **Immunoprecipitation (IP).**
594 For Immunoprecipitation (IP) and co-IP experiments, testes were dissected into 1X PBS on a
595 cyclops dissecting dish at room temperature from batches of 50 flies for a maximum of 30 minutes
596 per batch, snap-frozen in liquid nitrogen and stored at -80 °C. For one sample ~100 pairs of testes
597 were used. The day before the actual IP, 2 uL per sample of rabbit anti-HA antibodies (Cell
598 Signalling Technologies, #3724) were crosslinked to magnetic beads as described in Baker *et al.*
599 2023. The day of the experiment, beads crosslinked with antibody were rinsed 1 time in 200 mM
600 glycine pH 2.5 and washed once for 5 minutes in 200 mM glycine pH 2.5 to remove non-
601 crosslinked antibodies, then blocked for 1 hour in 10% BSA in 50 mM Tris pH 8, and re-suspended
602 in 20 uL of RIPA buffer and stored at 4 °C while cell lysate was prepared. Frozen tissue was
603 dissociated in RIPA buffer using a 1mL syringe and a 27G needle as in WB experiments. The lysate
604 was rocked at 4 °C for 20 minutes, then centrifuged 5 minutes at max speed in an Eppendorf tube

605 at 4 °C to separate lysate from tissue debris. Supernatant was transferred to a new Eppendorf
606 tube and was incubated 45 minutes at 4 °C with beads with no antibody. Tubes containing lysate
607 and beads were placed on a magnetic rack, 40 uL of lysate were saved as input sample and the
608 remaining volume (160 uL) was pipetted into the tubes containing beads crosslinked to antibody
609 and left incubating for 4 hours at 4 °C rocking, volume was brought back to 220 uL by adding 40
610 uL of cold RIPA buffer. Since we noticed that HA:Cbc in cell lysate tends to form aggregates (see
611 SOM Figure 11), we precipitated the input sample and resuspended in JUME Buffer. 4X Laemli
612 Sample Buffer (BIO-RAD #1610747) mixed with β-mercaptoethanol was added to the input
613 sample, incubated 7 minutes at 85 °C and frozen. After 4 hours rocking at 4 °C, IP samples were
614 placed on a magnet to separate beads from lysate, 40 uL of the supernatant were saved as
615 supernatant sample, mixed with 4X Laemli Sample Buffer with β-mercaptoethanol incubated 7
616 minutes at 85 °C and frozen; the rest of supernatant was tossed. The beads with attached protein
617 were immediately resuspended in RIPA buffer and washed with RIPA buffer 2 times for 5 minutes
618 rocking at room temperature. Tubes were placed on a magnet to separate beads from RIPA buffer,
619 then RIPA buffer was removed and beads were resuspended in 40 uL Elution Buffer (1X cOmplete
620 ULTRA Tablets, 1% SDS, 10 mM EDTA, 50 mM Tris pH 8), vortexed, and incubated for 30 minutes
621 at 70 °C with frequent (every 5 minutes) vortexing or shaking. Tubes were then centrifuged at
622 max speed for 10 seconds at room temperature, placed on the magnet to separate beads from
623 the eluate. 40 uL of eluate were moved to another tube (IP-sample) to which was added 4X Laemli
624 Sample Buffer with β-mercaptoethanol, incubated 7 minutes at 85 °C and frozen. The beads were
625 then mixed with 40 uL of elution buffer, mixed with 4X Laemli Sample Buffer with β-
626 mercaptoethanol, incubated 7 minutes at 85 °C and frozen. Samples were then run on a Mini-
627 PROTEAN TGX Precast Protein Gel and the protocol for WB above followed.
628 If the same membrane needed to be re-stained with another antibody, the membrane was
629 stripped by incubation at 55 °C for 30 minutes in 62.5 mM Tris pH 6.8, 2% SDS and 100 mM β-
630 mercaptoethanol, washed 3 times for 10 minutes each at room temperature in PBST and either
631 stored in 1X PBS or blocked again before being re-stained as above.
632
633

634 **Protein precipitation.**

635 4 volumes of ice-cold methanol are added to 1 volume of cell lysate, vortexed, and 1 volume of
636 chloroform added. The sample was vortexed again, mixed with 3 volumes of H_2O, vortexed and
637 centrifuged in 1.5 mL Eppendorf tubes at 4 ^{o}C max speed for 5 minutes. After centrifugation, 2
638 liquid layers formed, with proteins at the interface. Upper liquid layer was removed, and 3
639 volumes of ice-cold methanol added to the sample. After a quick vortex, the sample was
640 centrifuged at 4 ^{o}C max speed for 10 minutes to pellet the proteins. After centrifugation all liquid
641 was removed, and the white protein pellet was air-dried and resuspended in 40 uL of JUME buffer.
642

643 **Figure Legends**

644

645 **Figure 1. Expression of Cleavage Factors during *Drosophila* spermatogenesis**. (A) Schematic
646 representation of the cleavage machinery, adapted from Gallicchio *et al.* (2023). Grey: CPSF =
647 Cleavage and Polyadenylation Factor complex. Orange: CstF = Cleavage and Stimulation Factor
648 complex. Blue: CFI = Cleavage Factor I. Green: CFII = Cleavage Factor II. UE: Upstream element.
649 DE: Downstream element. PAS = Polyadenylation signal (B) Schematic representation of a WT
650 *Drosophila* testes showing key stages of germ cell differentiation. (C) Diagram of the
651 developmentally regulated APA in the *Drosophila* testis. (D) Quantification of relative
652 immunofluorescence signal per nuclear area from (E-J). Each dot corresponds to a single nucleus.
653 For each testis 20 spermatogonial nuclei and 20 spermatocyte nuclei were quantified, N=5 testes
654 were quantified per staining, for a total of 100 spermatogonia and 100 spermatocytes. Statistical
655 significance was measured using the Kolmogorov-Smirnov t-test: (****) p-value < 0.0001; (*) p =
656 0.0366. (E-J) Immunofluorescence images of testis apical regions. (Magenta) anti-Vasa marks
657 germ cells. (Cyan) cleavage factor proteins: (E-G, J) Wild type (*w1118*) testes stained with anti-
658 (E) CstF50, (F) CstF64, (G) PABP2 and (J) PCF11. (H) Testes from *CPSF6:GFP* transgenic flies stained
659 with anti-GFP, (I) Testes from *HA:cbc/HA:cbc* flies stained with anti-HA. Dotted yellow line: border
660 between proliferating spermatogonia (SG) and differentiating spermatocytes (SC). Scale bars = 50
661 uM.

662

663 **Figure 2. Knock down of *PCF11* and *cbc* in spermatocytes.** (A-B-C) Apical testes tip stained with
664 anti-PCF11. (A) Control: *bamGal4*, (B) *PCF11-RNAi-VDRC-103710;bamGal4* and (C) *PCF11-RNAi*
665 *NIG-10228R-4;bamGal4*. (D-E-F) Phase contrast microscopy images of whole mount live testes
666 from *Drosophila* lines in (A-B-C) respectively. (G-H-I) Apical testes tip stained with anti-HA. (G)
667 Control *HA:cbc/Cyo;bamGal4*, (H) *HA:cbc/+,cbc-RNAi-BL-77426;bamGal4* and (I) *HA:cbc/+,cbc-*
668 *RNAi-VDRC-100686;bamGal4*. (J-K-I) Phase contrast microscopy images of whole mount live
669 testes. (J) Control: *bamGal4*, (K) *cbc-RNAi-BL-77426;bamGal4* and (L) *cbc-RNAi-VDRC-*
670 *100686;bamGal4*. Arrows indicate somatic cysts cells. Scale bars are 50 uM (A-B-C-G-H-I) and 200
671 uM (D-E-F-J-K-L). Note that in all the immunofluorescence experiments on *HA:cbc/+* flies, the

level of immunofluorescence signal with anti-HA in spermatogonia was under the detection limit of the antibody, as the flies used in this experiment were heterozygous for the *HA:cbc/+*, while the testes quantified in Fig.1-I were homozygous *HA:cbc*/*HA:cbc* (for comparison between homozygous and heterozygous testes see SOM Figure 1).

Figure 3. Knockdown of either PCF11 or cbc reduces cleavage of the proximal PAS in spermatocytes. (A) Schematic representation of a gene body (either *Kpc2* or *Red*) to illustrate design of the two probe sets. (B-C-D, B''-C''-D'') Confocal microscopy images of (B, B'') *bamGal4* (control), (C, C'') *PCF11-RNAi-VDRC-103710;bamGal4* and (D, D'') *cbc-RNAi-BL-77426;bamGal4* testes stained against probes recognizing the CDS of the gene *Kpc2* (in cyan) and DAPI (B-C-D), and against probes recognizing the 3'UTR extension of the gene *Kpc2* (in orange) (B''-C''-D''). (B'-C'-D', B'''-C'''-D''') zooms in spermatocytes of corresponding regions from (B-C-D, B''-C''-D'') respectively. (E) 3'seq pile reads for the gene *Kpc2* from a sample enriched for SG = spermatogonia (16H PHS = post heat shock, Kim *et al.* 2017; Berry *et al.* 2022) and a sample enriched for spermatocytes SC = spermatocytes (72H PHS) to show the APA switch from the distal PAS (dPAS) used in spermatogonia and the proximal PAS (pPAS) used in spermatocytes. (F) Quantification of fluorescence ratio between fluorescence coming from the probes recognizing the *Kpc2* 3'UTR extension (orange) and fluorescence recognizing the *Kpc2* CDS (cyan) in spermatocytes. For each genotype N=3 testes were quantified. (G-K) Analogous to (A-F) for the gene *Red*. Scale bars are 150 uM (B-B''-C-C''-D-D''-G-G''-H-H''-I-I'') and 50 uM (B'-B'''-C'-C'''-D'-D'''-G'-G'''-H'-H'''-I'-I'''). p values are (****) = $p<0.0001$ and (***) = $p = 0.0002$. For statistical analysis we used the Welch's t-test.

Figure 4. 3' end RNAseq libraries comparing effects on APA of knockdowns in spermatocytes of different components of the cleavage machinery. (A-B) Integrative Genomics Viewer (IGV) plots of 3' end RNA sequencing libraries for the genes (A) *Kpc2* and (B) *Red*. Regions of peaks that change in color show mismatches between the PolyA tail and the genome to highlight the 3' end of the mRNA. (C) Heatmap of PolyA index quantified with PolyA Miner across 3'seq libraries. Each row represents one of the >500 genes that switch to a proximal PAS in WT spermatocytes

701 compared to spermatogonia from Berry *et al.* (2022). Each column represents a component of
702 the cleavage machinery that was KD in spermatocytes. The interval chosen to plot the PolyA index
703 (-1.5,1.5) does not represent the maximum and minimum PolyA index observed. In this heatmap
704 genes that showed a statistically non-significant PolyA index in all the samples or a negative PolyA
705 index in all the samples were removed, and in all other cases PolyA indexes that were non-
706 significant were put equal to zero (another heatmap with these genes showed is reported in SOM
707 Figure 2) (D-E) Integrative Genomics Viewer (IGV) plots of 3' end RNA sequencing libraries for the
708 genes (D) *oaf* and (E) *nudE*. (F-G-H) Venn diagrams representing number of genes whose polyA
709 index is > 0.5 in *cbc* KD libraries (dark green), *PCF11* VDRC KD libraries (light green), *PABP2* KD
710 libraries (red) and components of CstF complex (orange).
711
712 **Figure 5. *PCF11* and *cbc* transcripts are both up-regulated in spermatocytes with *cbc* particularly**
713 **highly expressed in testis compared to other tissues.** (A) Diagram adapted from FlyBase of *PCF11*
714 and *cbc* loci. Light blue boxes represent 5'UTRs, grey boxes represent exons and pink boxes
715 represent 3'UTRs, while introns are represented with a straight black line. In yellow are
716 represented the transcripts regions against which we designed HCR-FISH probes for RNA *in situ*.
717 Green and dark green boxes on *PCF11* locus represent the CID domain and a predicted Cbc
718 binding domain respectively. (B-G) Scope diagrams of Fly Cell Atlas single nuclear RNAseq (Li *et*
719 *al.* 2023) data showing: (B) *PCF11* and (C) *cbc* expression across all the analyzed tissues, (D) Umap
720 highlighting the testis cluster across all analyzed tissues (in purple), (E) *PCF11* and (F) *cbc*
721 expression in the testis sample, (G) Umap highlight germ-line cell clusters in the testis sample (in
722 orange). (H-H'-H'') Confocal microscopy of WT (*w1118*) testes stained with probes against *PCF11*
723 long isoforms (white) and DAPI (blue). (I-I'-I'') Confocal microscopy of WT (*w1118*) testes stained
724 with probes against *PCF11*-RB isoform (white) and DAPI (blue). (J-J'-J'') Confocal microscopy of
725 WT (*w1118*) testes stained with probes against *cbc* (white) and DAPI (blue). In (H-I-J) the boxes
726 highlight a region of spermatogonia zoomed in (H'-I'-J') and a region of spermatocytes zoomed
727 in (H''-I''-J''). (K-L) Umaps from Fly Cell Atlas single nuclear data reporting (K) *PCF11-long* and (L)
728 *PCF11-RB* expression across the different clusters of the testis plus seminal vesicle sample.
729 Scalebars are 50 uM in (H-I-J) and 20 uM in (H'-H''-I'-I''-J'-J'').

730

731 **Figure 6. *PCF11* KD causes loss of Cbc**. (A) Western Blots of WT (*w1118*) testis lysate stained
732 against endogenous PCF11. (B) Western Blots of *HA:cbc* homozygous testes lysate stained against
733 HA. (C) Immunoprecipitation (IP) of *HA:cbc* in WT (*w1118*) and *HA:cbc* homozygous testes
734 followed by western blots stained against endogenous PCF11 and against HA. Input = 20% of
735 lysate before IP. (D-E) IGV plots of (D) *cbc* and (E) *PCF11* loci of RNAseq data from *bamGal4*
736 (control) samples (in grey), *PCF11-RNAi-VDRC-103710;bamGal4* samples (magenta) and *cbc-*
737 *RNAi-BL-77426;bamGal4* samples (teal). (F-J) Immunofluorescence images showing apical tips of
738 testes stained against Vasa to mark germ cells, HA and PCF11. (F-F'-F'') *HA:cbc/cyo;bamGal4*
739 (control), (G-G'-G'') *HA:cbc/+,PCF11-RNAi-VDRC-103710;bamGal4*, (H-H'-H'') *HA:cbc/+,PCF11-*
740 *RNAi-NIG-10228R-4;bamGal4*, (I-I'-I'') *HA:cbc/+,cbc-RNAi-BL-77426;bamGal4* and (J-J'-J'')
741 *HA:cbc/+,cbc-RNAi-VDRC-100686;bamGal4.* (K-L) Quantification of Fluorescence/Area coming
742 from HA antibodies (teal) and from PCF11 antibodies (magenta). For each testis 10 spermatogonia
743 and 10 spermatocytes nuclei were quantified simultaneously for PCF11 and HA fluorescence and
744 then averaged. For each genotype N=3 testes were quantified. Statistical analysis was performed
745 using the Welch's t-test. p-values are: *bamGal4_PCF11* = 0.0011 (**), b*amGal4_cbc* = 0.0005
746 (***), *PCF11-RNAi-NIG10228R-4_PCF11* = 0.0341 (*), *cbc-RNAi-BL77427_PCF11* < 0.0001 (****),
747 *cbc-RNAi-VDRC100686_PCF11* = 0.0303 (*). Scalebars are 50 uM (F-J''). Entire gels for western
748 blots and co-IP are reported in SOM Figure 11.

Acknowledgements

We thank all the members of the Fuller lab for suggestions and comments on this manuscript and Dr. Nicole Martinez and Dr. Lauren Goins for their valuable feedback on the manuscript and suggested experiments. We thank Dr. Catherine Baker and Benjamin Bolival for their help in maintaining fly crosses. We thank Dr. Jingxun Chen from the Brunet lab at Stanford for her help in designing HCR-FISH probes and for sharing some of the reagents, the Wang lab for the *HA:cbc* CRISPR tag flies, the Lis lab for anti-CstF50, the Simonelig lab for anti-PABP2, the Dominski lab for anti-CstF64 and the Gilmour lab for anti-PCF11 antibodies. We thank Daniel Mokhtari for his help parsing the Fly Cell Atlas dataset. We thank the Cell Science Imaging facility at Stanford, supported, in part, by Award Number 1S10OD010580-01A1 from the National Center for Research Resources (NCRR) (The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the NCRR or the National Institutes of Health), particularly the CSIF facility manager Kitty Lee. We thank the Stanford Fly Media Center as well and all the staff working in the Developmental Biology department. L.G. was supported by an AICF (American Italian Cancer Foundation) postdoctoral fellowship (2021-2023). This research was supported by National Institutes of Health grant R35 GM136433) and funds from the Katherine D. McCormick and Stanley McCormick Memorial Professorship and the Reed-Hodgson Professorship in Human Biology to M.T.F. This manuscript is dedicated to Cuchuflí.

Author Contributions

LG and MTF conceived the project and wrote the manuscript. LG designed the experimental work and performed or supervised all experimental and computational procedures. NRM contributed to experimental design and bioinformatic analysis. FMP contributed to experimental design and performed molecular cloning. IN performed immunofluorescence experiments. SS performed bioinformatic analysis of the Fly Cell Atlas data. YZ performed bioinformatic analysis to perform *de novo* mapping of PASs. Funding was acquired by LG and MTF.

775 **References**

776 An, J. J., K. Gharami, G. Y. Liao, N. H. Woo, A. G. Lau *et al.*, 2008 Distinct Role of Long 3' UTR
777 BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. Cell.
778 Baker, C. C., L. Gallicchio, N. R. Matias, D. F. Porter, L. Parsanian *et al.*, 2023 Cell-type-specific
779 interacting proteins collaborate to regulate the timing of Cyclin B protein expression in
780 male meiotic prophase. Development dev.201709.
781 Bedbrook, C. N., R. D. Nath, R. Nagvekar, K. Deisseroth, and A. Brunet, 2023 Rapid and precise
782 genome engineering in a naturally short-lived vertebrate (L. Moons, D. Y. R. Stainier, L.
783 Moons, & P. Abitua, Eds.). Elife 12: e80639.
784 Benoit, B., M. Simonelig, A. Nemeth, N. Aulner, U. Kühn *et al.*, 1999 The Drosophila poly(A)-
785 binding protein II is ubiquitous throughout Drosophila development and has the same
786 function in mRNA polyadenylation as its bovine homolog in vitro. Nucleic Acids Res. 27:
787 3771–3778.
788 Berry, C. W., G. H. Olivares, L. Gallicchio, G. Ramaswami, A. Glavic *et al.*, 2022 Developmentally
789 regulated alternate 3' end cleavage of nascent transcripts controls dynamic changes in
790 protein expression in an adult stem cell lineage. Genes Dev. 36: 916–935.
791 Chen, D., and D. M. McKearin, 2003 A discrete transcriptional silencer in the bam gene
792 determines asymmetric division of the Drosophila germline stem cell. Development 130:
793 1159–1170.
794 Clerici, M., M. Faini, L. M. Muckenfuss, R. Aebersold, and M. Jinek, 2018 Structural basis of
795 AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat. Struct. Mol.
796 Biol. 25: 135–138.
797 Gallicchio, L., G. H. Olivares, C. W. Berry, and M. T. Fuller, 2023 Regulation and function of
798 alternative polyadenylation in development and differentiation. RNA Biol. 20: 908–925.
799 Grassi, E., R. Santoro, A. Umbach, A. Grosso, S. Oliviero *et al.*, 2019 Choice of alternative
800 polyadenylation sites, mediated by the rna-binding protein Elavl3, plays a role in
801 differentiation of inhibitory neuronal progenitors. Front. Cell. Neurosci.
802 Grozdanov, P. N., E. Masoumzadeh, M. P. Latham, and C. C. MacDonald, 2018 The structural
803 basis of CstF-77 modulation of cleavage and polyadenylation through stimulation of CstF-

804 64 activity. Nucleic Acids Res. 46: 12022–12039.
805 Gruber, A. J., and M. Zavolan, 2019 Alternative cleavage and polyadenylation in health and
806 disease. Nat. Rev. Genet.
807 Guéguéniat, J., A. F. Dupin, J. Stojko, L. Beaurepaire, S. Cianférani *et al.*, 2017 Distinct roles of
808 PCF11 zinc-binding domains in pre-mRNA 3-end processing. Nucleic Acids Res.
809 Hilgers, V., S. B. Lemke, and M. Levine, 2012 ELAV mediates 3' UTR extension in the Drosophila
810 nervous system. Genes Dev. 26: 2259–2264.
811 Jing, Z., H. Linda, and M. Claire, 1999 Formation of mRNA 3' Ends in Eukaryotes: Mechanism,
812 Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol. Mol.
813 Biol. Rev. 63: 405–445.
814 Jongmin Kim, Chenggang Lu, Shrividhya Srinivasan, Stephan Awe, Alexander Brehm *et al.*, 2017
815 Blocking promiscuous activation at cryptic promoters directs cell type – specific gene
816 expression. Science (80-.).
817 Kamieniarz-Gdula, K., M. R. Gdula, K. Panser, T. Nojima, J. Monks *et al.*, 2019 Selective Roles of
818 Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Mol. Cell 74: 158-
819 172.e9.
820 Lee, S., Y.-C. Chen, A. E. Gillen, J. M. Taliaferro, B. Deplancke *et al.*, 2022 Diverse cell-specific
821 patterns of alternative polyadenylation in Drosophila. Nat. Commun. 13: 5372.
822 Li, H., J. Janssens, M. De Waegeneer, S. S. Kolluru, K. Davie *et al.*, 2023 Fly Cell Atlas: A single-
823 nucleus transcriptomic atlas of the adult fruit fly. Science (80-.). 375: eabk2432.
824 Li, W., J. Y. Park, D. Zheng, M. Hoque, G. Yehia *et al.*, 2016 Alternative cleavage and
825 polyadenylation in spermatogenesis connects chromatin regulation with post-
826 transcriptional control. BMC Biol.
827 Li, W., B. You, M. Hoque, D. Zheng, W. Luo *et al.*, 2015 Systematic Profiling of Poly(A)+
828 Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory
829 Rules of Alternative Cleavage and Polyadenylation. PLOS Genet. 11: e1005166.
830 Mansfield, K. D., and J. D. Keene, 2012 Neuron-specific ELAV/Hu proteins suppress HuR mRNA
831 during neuronal differentiation by alternative polyadenylation. Nucleic Acids Res. 40:
832 2734–2746.

833 Marzluff, W. F., E. J. Wagner, and R. J. Duronio, 2008 Metabolism and regulation of canonical
834 histone mRNAs: Life without a poly(A) tail. Nat. Rev. Genet.
835 Mayr, C., and D. P. Bartel, 2009 Widespread Shortening of 3'UTRs by Alternative Cleavage and
836 Polyadenylation Activates Oncogenes in Cancer Cells. Cell 138: 673–684.
837 Miura, P., S. Shenker, C. Andreu-Agullo, J. O. Westholm, and E. C. Lai, 2013 Widespread and
838 extensive lengthening of 39 UTRs in the mammalian brain. Genome Res.
839 Morris, A. R., A. Bos, B. Diosdado, K. Rooijers, R. Elkon *et al.*, 2012 Alternative cleavage and
840 polyadenylation during colorectal cancer development. Clin. Cancer Res. 18: 5256–5266.
841 Ni, Z., A. Saunders, N. J. Fuda, J. Yao, J.-R. Suarez *et al.*, 2008 P-TEFb Is Critical for the
842 Maturation of RNA Polymerase II into Productive Elongation In Vivo. Mol. Cell. Biol. 28:
843 1161–1170.
844 Ogorodnikov, A., M. Levin, S. Tattikota, S. Tokalov, M. Hoque *et al.*, 2018 Transcriptome 3'end
845 organization by PCF11 links alternative polyadenylation to formation and neuronal
846 differentiation of neuroblastoma. Nat. Commun.
847 Raz, A. A., G. S. Vida, S. R. Stern, S. Mahadevaraju, J. M. Fingerhut *et al.*, 2023 Emergent
848 dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of
849 Drosophila testes (M. Buszczak, U. Banerjee, M. Buszczak, & T. G. Nystul, Eds.). Elife 12:
850 e82201.
851 Sadowski, M., B. Dichtl, W. Hübner, and W. Keller, 2003 Independent functions of yeast Pcf11p
852 in pre-mRNA 3' end processing and in transcription termination. EMBO J. 22: 2167–2177.
853 Sarov, M., C. Barz, H. Jambor, M. Y. Hein, C. Schmied *et al.*, 2016 A genome-wide resource for
854 the analysis of protein localisation in Drosophila. Elife 5:.
855 Schäfer, P., C. Tüting, L. Schönemann, U. Kühn, T. Treiber *et al.*, 2018 Reconstitution of
856 mammalian cleavage factor II involved in 3' processing of mRNA precursors. RNA.
857 Shan, L., C. Wu, D. Chen, L. Hou, X. Li *et al.*, 2017 Regulators of alternative polyadenylation
858 operate at the transition from mitosis to meiosis. J. Genet. Genomics 44: 95–106.
859 Skrajna, A., X. Yang, M. Dadlez, W. F. Marzluff, and Z. Dominski, 2018 Protein composition of
860 catalytically active U7-dependent processing complexes assembled on histone pre-mRNA
861 containing biotin and a photo-cleavable linker. Nucleic Acids Res. 46: 4752–4770.

862 Tian, B., and J. L. Manley, 2016 Alternative polyadenylation of mRNA precursors. Nat. Rev. Mol.
863 Cell Biol.
864 Turner, R. E., L. M. Henneken, M. Liem-Weits, P. F. Harrison, A. Swaminathan *et al.*, 2020
865 Requirement for cleavage factor IIm in the control of alternative polyadenylation in breast
866 cancer cells. RNA 26: 969–981.
867 Vallejos Baier, R., J. Picao-Osorio, and C. R. Alonso, 2017 Molecular Regulation of Alternative
868 Polyadenylation (APA) within the Drosophila Nervous System. J. Mol. Biol.
869 Venkat, S., A. A. Tisdale, J. R. Schwarz, A. A. Alahmari, H. Carlo Maurer *et al.*, 2020 Alternative
870 polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma.
871 Genome Res. 30: 347–360.
872 Wu, J., X. Li, Z. Gao, L. Pang, X. Liu *et al.*, 2021 RNA kinase CLP1/Cbc regulates meiosis initiation
873 in spermatogenesis. Hum. Mol. Genet. 30: 1569–1578.
874 Yalamanchili, H. K., C. E. Alcott, P. Ji, E. J. Wagner, H. Y. Zoghbi *et al.*, 2020 PolyA-miner:
875 Accurate assessment of differential alternative poly-adenylation from 3'Seq data using
876 vector projections and non-negative matrix factorization. Nucleic Acids Res.
877 Zeng, Y., A. Lovchykova, T. Akiyama, C. Liu, C. Guo *et al.*, 2024 TDP-43 nuclear loss in FTD/ALS
878 causes widespread alternative polyadenylation changes. bioRxiv 2024.01.22.575730.
879 Zhang, Z., and D. S. Gilmour, 2006 Pcf11 Is a Termination Factor in Drosophila that Dismantles
880 the Elongation Complex by Bridging the CTD of RNA Polymerase II to the Nascent
881 Transcript. Mol. Cell 21: 65–74.
882 Zhu, Y., X. Wang, E. Forouzmand, J. Jeong, F. Qiao *et al.*, 2018 Molecular Mechanisms for CFIm-
883 Mediated Regulation of mRNA Alternative Polyadenylation. Mol. Cell 69: 62-74.e4.
884
885
886

Fig.1

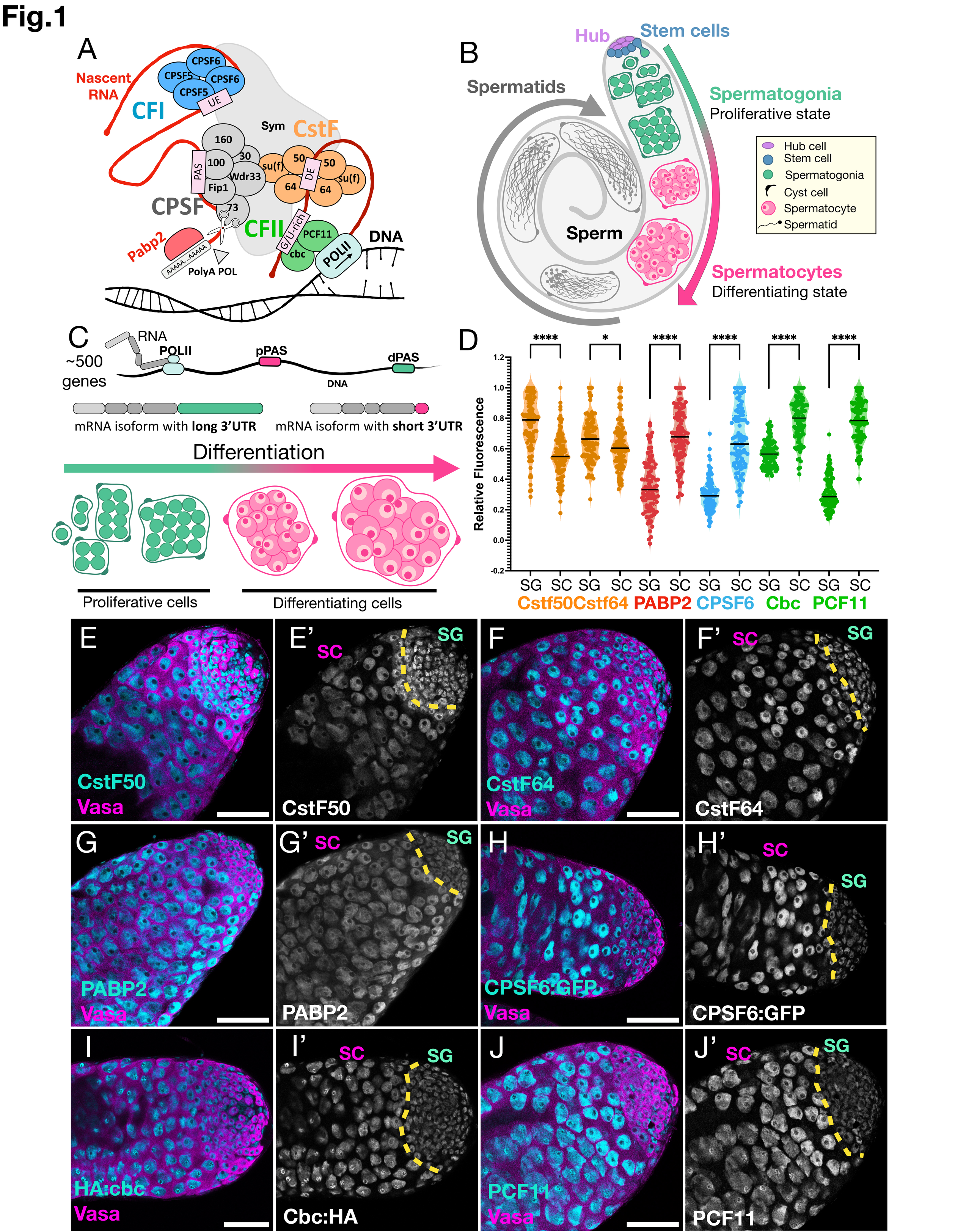

A
Nascent RNA
CPSF6
CPSF5
CPSF6
CPSF5
UE
CFI
Sym
CstF
160
30
100
Wdr33
Fip1
73
PAS
su(f)
50
64
DE
50
64
su(f)
CPSF
CFII
G/U-rich
PCF11
cbc
POLII
DNA
Pabp2
AAAAA...AAAAA
PolyA POL
B
Hub
Stem cells
Spermatids
Spermatogonia
Proliferative state
Hub cell
Stem cell
Spermatogonia
Cyst cell
Spermatocyte
Spermatid
Sperm
Spermatocytes
Differentiating state
C
~500 genes
RNA
POLII
pPAS
dPAS
DNA
mRNA isoform with long 3'UTR
mRNA isoform with short 3'UTR
Differentiation
Proliferative cells
Differentiating cells
D
Relative Fluorescence
SG SC SG SC SG SC SG SC SG SC SG SC
Cstf50Cstf64 PABP2 CPSF6 Cbc PCF11
E
CstF50
Vasa
E'
SC
SG
CstF50
F
CstF64
Vasa
F'
SC
SG
CstF64
G
PABP2
Vasa
G'
SC
SG
PABP2
H
CPSF6:GFP
Vasa
H'
SC
SG
CPSF6:GFP
I
HA:cbc
Vasa
I'
SC
SG
Cbc:HA
J
PCF11
Vasa
J'
SC
SG
PCF11

Fig.2

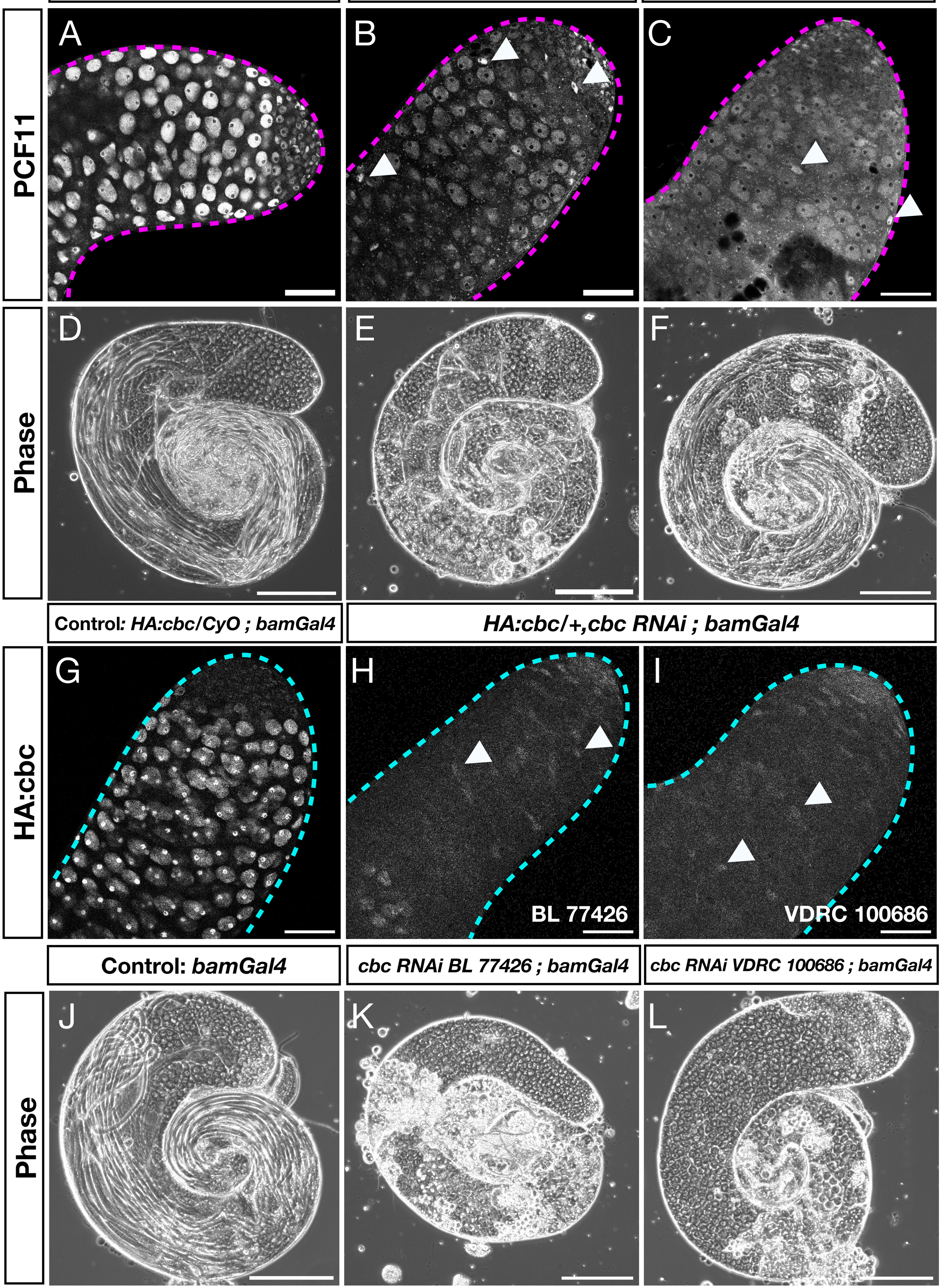

Control: bamGal4
PCF11 RNAi VDRC 103710 ; bamGal4
PCF11 RNAi NIG 10228R-4 ; bamGal4
PCF11
A
B
C
Phase
D
E
F
Control: HA:cbc/CyO ; bamGal4
HA:cbc/+,cbc RNAi ; bamGal4
HA:cbc
G
H
I
BL 77426
VDRC 100686
Control: bamGal4
cbc RNAi BL 77426 ; bamGal4
cbc RNAi VDRC 100686 ; bamGal4
Phase
J
K
L

Fig.3

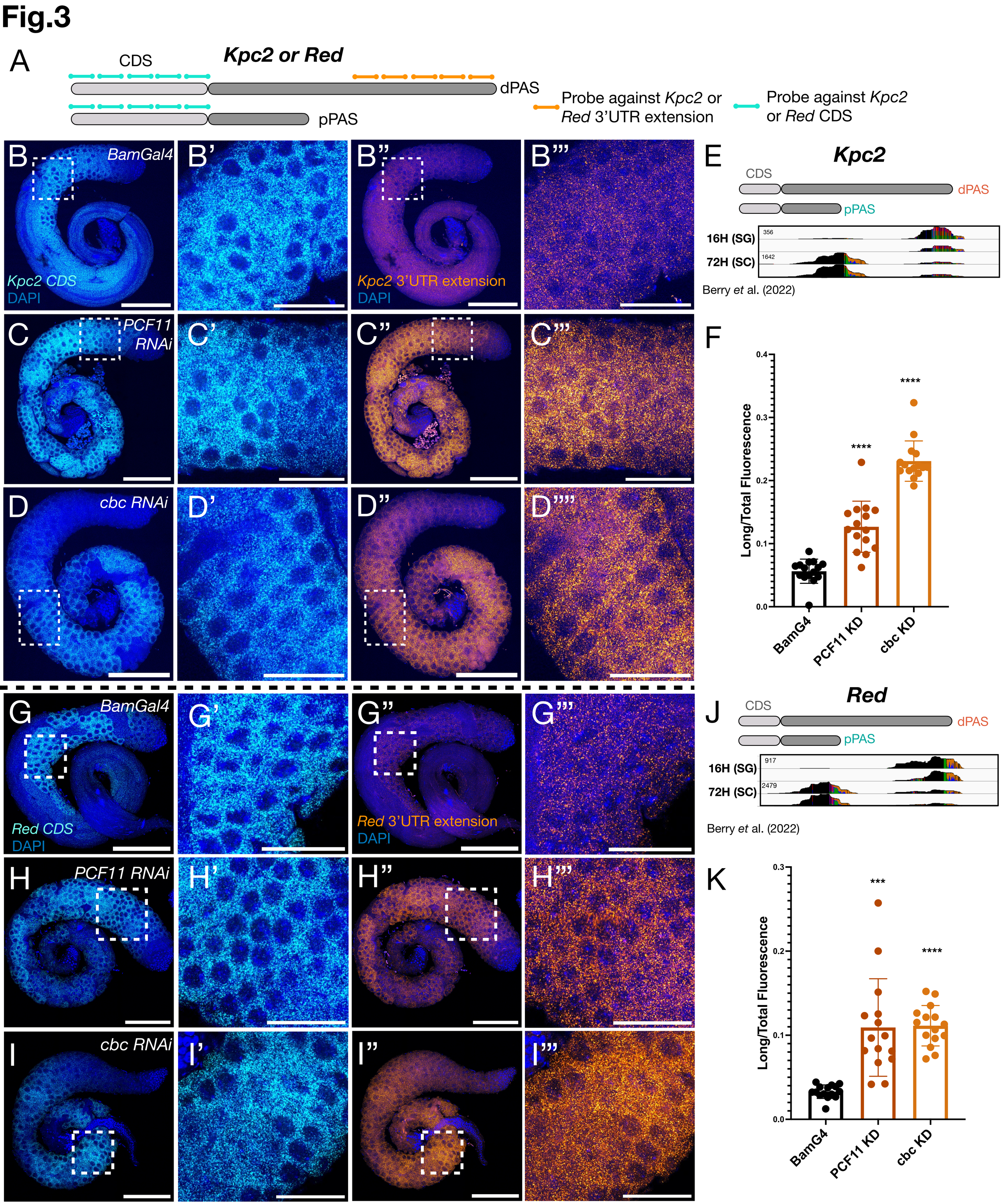

A
CDS
Kpc2 or Red
dPAS
pPAS
Probe against Kpc2 or Red 3'UTR extension
Probe against Kpc2 or Red CDS
B
BamGal4
Kpc2 CDS
DAPI
B'
B''
Kpc2 3'UTR extension
DAPI
B'''
C
PCF11 RNAi
C'
C''
C'''
D
cbc RNAi
D'
D''
D''''
E
Kpc2
CDS
dPAS
pPAS
16H (SG)
72H (SC)
356
1642
Berry et al. (2022)
F
Long/Total Fluorescence
0.0
0.1
0.2
0.3
0.4
****
****
BamG4
PCF11 KD
cbc KD
G
BamGal4
Red CDS
DAPI
G'
G''
Red 3'UTR extension
DAPI
G'''
H
PCF11 RNAi
H'
H''
H'''
I
cbc RNAi
I'
I''
I'''
J
Red
CDS
dPAS
pPAS
16H (SG)
72H (SC)
917
2479
Berry et al. (2022)
K
Long/Total Fluorescence
0.0
0.1
0.2
0.3
***
****
BamG4
PCF11 KD
cbc KD

Fig.4

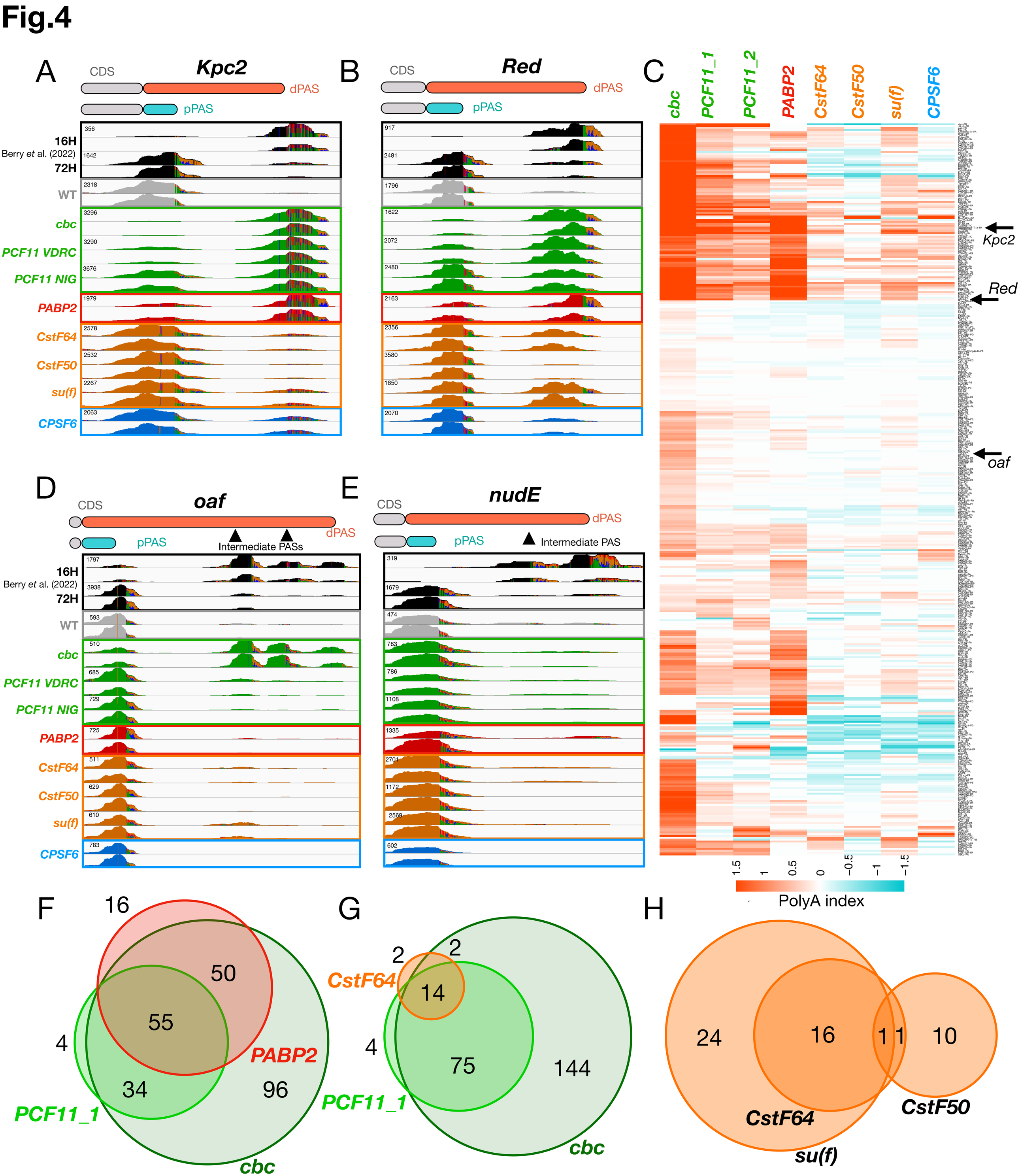

A
Kpc2
CDS
dPAS
pPAS
16H
Berry et al. (2022)
72H
WT
cbc
PCF11 VDRC
PCF11 NIG
PABP2
CstF64
CstF50
su(f)
CPSF6
B
Red
C
cbc
PCF11_1
PCF11_2
PABP2
CstF64
CstF50
su(f)
CPSF6
Kpc2
Red
oaf
PolyA index
D
oaf
Intermediate PASs
E
nudE
Intermediate PAS
F
16
50
55
4
34
96
PABP2
PCF11_1
cbc
G
2
2
CstF64
14
4
75
144
PCF11_1
cbc
H
24
16
1
1
10
CstF64
CstF50
su(f)

Fig.5

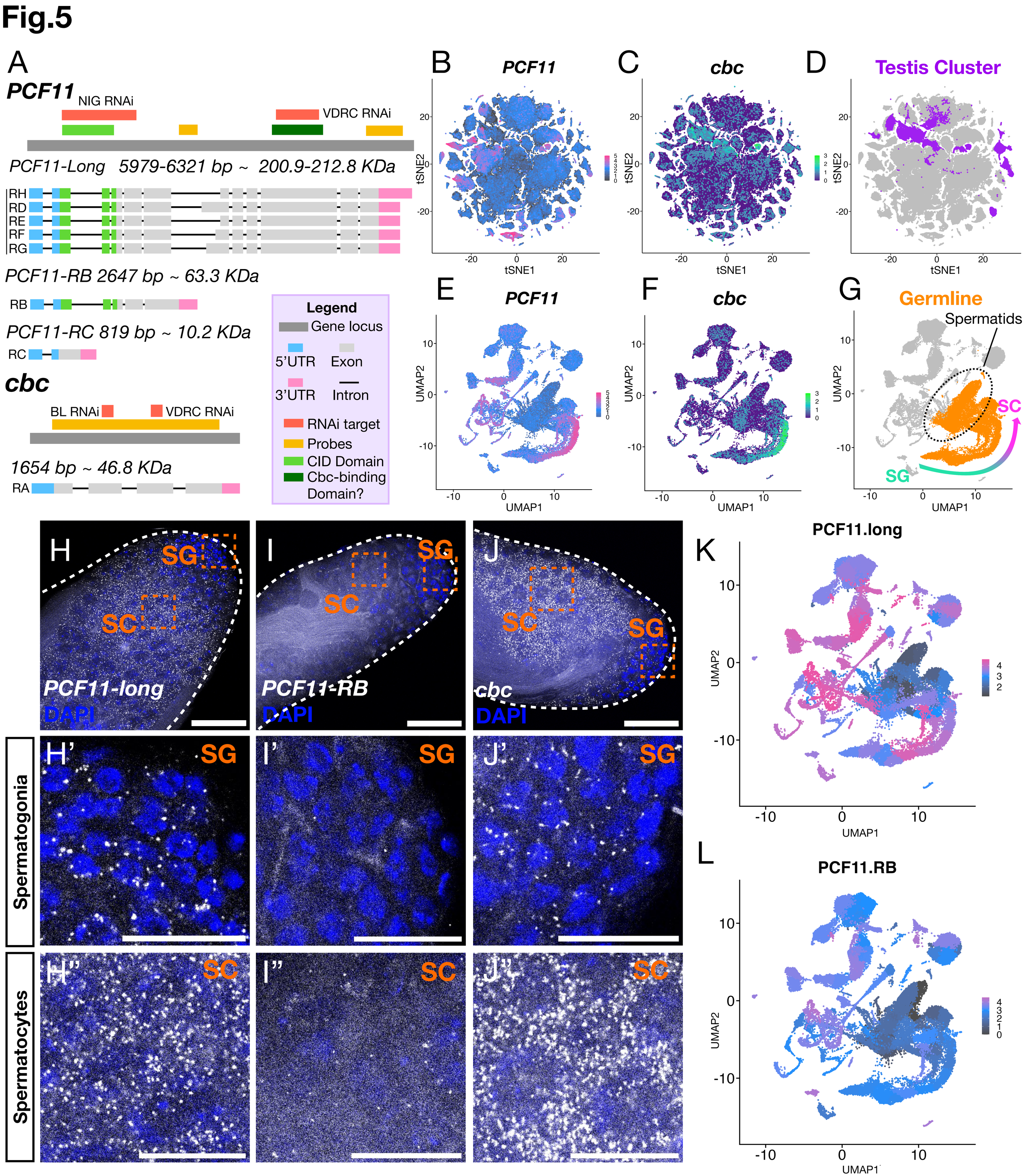
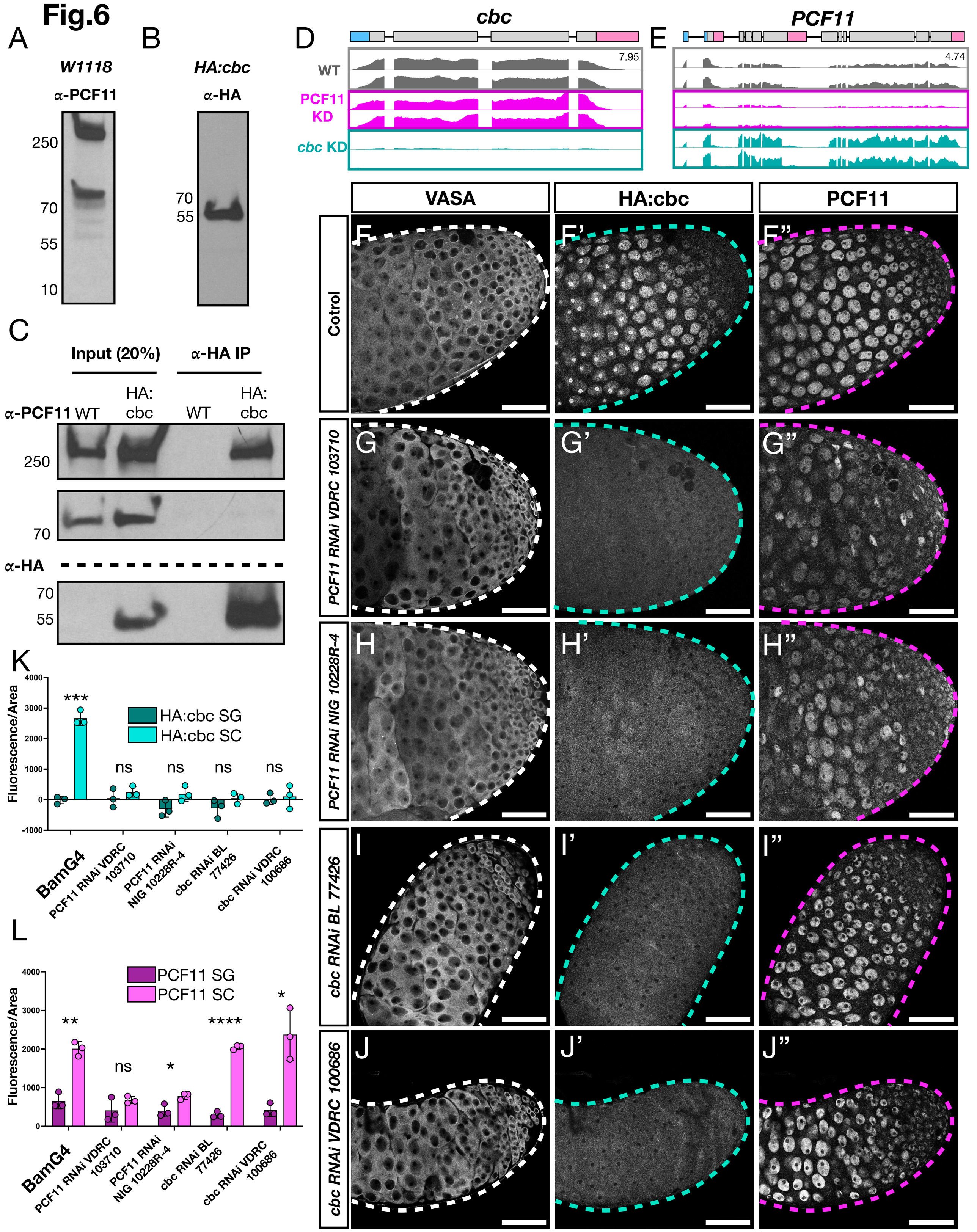

A
PCF11
NIG RNAi
VDRC RNAi
PCF11-Long 5979-6321 bp ~ 200.9-212.8 KDa
RH
RD
RE
RF
RG
PCF11-RB 2647 bp ~ 63.3 KDa
RB
PCF11-RC 819 bp ~ 10.2 KDa
RC
cbc
BL RNAi
VDRC RNAi
1654 bp ~ 46.8 KDa
RA
Legend
Gene locus
5'UTR
Exon
3'UTR
Intron
RNAi target
Probes
CID Domain
Cbc-binding Domain?
B
PCF11
C
cbc
D
Testis Cluster
tSNE1
tSNE2
E
PCF11
F
cbc
G
Germline
Spermatids
SC
SG
UMAP1
UMAP2
H
I
J
SG
SC
PCF11-long
PCF11-RB
cbc
DAPI
H'
I'
J'
H''
I''
J''
Spermatogonia
Spermatocytes
K
PCF11.long
L
PCF11.RB

Fig.6


A
W1118
α-PCF11
250
70
55
10
B
HA:cbc
α-HA
70
55
C
Input (20%)
α-HA IP
α-PCF11
WT
HA: cbc
WT
HA: cbc
250
70
α-HA
70
55
D
cbc
WT
PCF11 KD
cbc KD
7.95
E
PCF11
4.74
VASA
HA:cbc
PCF11
Cotrol
PCF11 RNAi VDRC 103710
PCF11 RNAi NIG 10228R-4
cbc RNAi BL 77426
cbc RNAi VDRC 100686
F
F'
F''
G
G'
G''
H
H'
H''
I
I'
I''
J
J'
J''
K
Fluorescence/Area
4000
3000
2000
1000
0
-1000
***
ns
ns
ns
ns
HA:cbc SG
HA:cbc SC
BamG4
PCF11 RNAi VDRC 103710
PCF11 RNAi NIG 10228R-4
cbc RNAi BL 77426
cbc RNAi VDRC 100686
L
Fluorescence/Area
4000
3000
2000
1000
0
PCF11 SG
PCF11 SC
**
ns
*
****
*
BamG4
PCF11 RNAi VDRC 103710
PCF11 RNAi NIG 10228R-4
cbc RNAi BL 77426
cbc RNAi VDRC 100686