A Developmental Mechanism to Regulate 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, *Drosophila*, mRNA processing, Cleavage factor Complex II.

Abstract

Alternative Cleavage and Polyadenylation (APA) often results in production of mRNA isoforms with either longer or shorter 3'UTRs from the same genetic locus, potentially impacting mRNA translation, localization and stability. Developmentally regulated APA can thus make major contributions to cell-type-specific gene expression programs as cells differentiate. During *Drosophila* spermatogenesis, approximately 500 genes undergo APA when proliferating spermatogonia differentiate into spermatocytes, producing transcripts with shortened 3' UTRs, leading to 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 upregulation of PCF11 and Cbc, the two components of Cleavage Factor II (CFII), orchestrates APA during *Drosophila* spermatogenesis. Knock down 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. Forced overexpression of CFII components in spermatogonia switched cleavage of some transcripts to the proximal site normally used in spermatocytes. Our findings reveal a developmental mechanism where changes in expression of specific cleavage factors can direct cell-type-specific APA at selected genes.

38 **Introduction**

39 Alternative cleavage and polyadenylation (APA) of mRNAs results from changes in the
40 site at which the 3' end cut that terminates nascent transcripts is made (Tian and Manley 2016;
41 Gallicchio *et al.* 2023). For alternative polyadenylation sites located in the region of a gene
42 encoding the 3'UTR, APA results in production of alternate mRNA isoforms that differ in 3'UTR
43 length and content, with potentially profound effects on transcript fate, including translation,
44 stability, and localization (An *et al.* 2008; Mayr and Bartel 2009; Mansfield and Keene 2012;
45 Berry *et al.* 2022). More than 70% of human genes are subject to alternative polyadenylation,
46 making APA a widespread biological phenomenon (Gruber and Zavolan 2019). The most
47 intriguing APA instances are those where the choice of 3' end cut site is developmentally
48 regulated, leading, for example, to production of transcripts with longer 3'UTRs from specific
49 genes in neurons or shorter 3'UTRs in differentiating male germ cells compared to precursor
50 cells (Hilgers *et al.* 2012; Miura *et al.* 2013; Shan *et al.* 2017; Vallejos Baier *et al.* 2017; Grassi *et*
51 *al.* 2019; Berry *et al.* 2022). Changes in 3'UTR length due to APA also occur in certain
52 pathological states, such as the 3'UTR shortening of transcripts observed in some cancers (Mayr
53 and Bartel 2009; Morris *et al.* 2012; Venkat *et al.* 2020). Because these changes in 3'UTR
54 content can have dramatic effects on protein expression and thus cell state, cell-type-specific
55 regulation of APA is emerging as an important developmental mechanism to promote robust
56 cell fate decisions.

57 During cleavage and polyadenylation, the last step of mRNA synthesis, the nascent
58 mRNA molecule is cleaved and a string of adenosines is added by nuclear PolyA polymerase to
59 protect the newly made transcript from degradation and facilitate its export from the nucleus
60 (Jing *et al.* 1999). Most mRNAs undergo cleavage and polyadenylation, with a notable exception
61 being histone transcripts, which are not polyadenylated (Marzluff *et al.* 2008). Four well-
62 conserved multi-protein complexes have been identified as components of the Cleavage and
63 Polyadenylation machinery (Figure 1 A). The core complex CPSF (Cleavage and Polyadenylation
64 Specificity Factor) is responsible for recognizing (via the CPSF30 and WDR33 subunits) the
65 polyadenylation signal (PAS) and inducing cleavage of the nascent RNA molecule (via its CPSF73
66 subunit). The PAS is an hexamer (most commonly AAUAAA or a close variant) in the nascent

67 mRNA usually ~15-30 nucleotides upstream of the cleavage site (Clerici *et al.* 2018; Gallicchio *et*
68 *al.* 2023). Another complex, CstF (Cleavage and Stimulation Factor), recognizes U/GU-rich
69 regions downstream of the PAS (MacDonald *et al.* 1994). CFI (Cleavage Factor I) recognizes a
70 UGUA motif upstream of the PAS (Yang *et al.* 2011) and CFII (Cleavage Factor II) preferentially
71 binds to G-rich sequences via its PCF11 subunit (Schäfer *et al.* 2018). The Cleavage and
72 Polyadenylation machinery is also associated with several additional proteins: Symplekin (a
73 scaffolding protein), PABP2, and the nuclear PolyA polymerase (Figure 1 A) (Gallicchio *et al.*
74 2023). While the central role of the core complex CPSF in recognizing the PAS and cleaving the
75 nascent mRNA molecule is relatively clear, the specific contributions of the other complexes and
76 associated factors in specifying whether, where, on what specific gene products, and in which
77 cell types a cut is made are still under investigation (Vallejos Baier *et al.* 2017; Zhu *et al.* 2018;
78 Turner *et al.* 2020).

79 Developmentally regulated changes in the PAS site utilized can have profound effects
80 on the suite of proteins expressed as cells progress from one state to the next in a
81 developmental sequence. For example, during spermatogenesis in flies and mice, many genes
82 express mRNA isoforms with long 3'UTRs in proliferating spermatogonia, but isoforms with
83 shorter 3'UTRs at later stages of germ cell development (Li *et al.,* 2016; Lee *et al.,* 2022; Berry *et*
84 *al.,* 2022). In *Drosophila melanogaster* testes (Figure 1 B), over 500 genes express transcripts
85 cleaved at a distal PAS in proliferating spermatogonia but at a more proximal PAS in
86 differentiating spermatocytes, resulting in production of mRNA isoforms with much shorter
87 3'UTRs (Figure 1 C).

88 The switch from distal to proximal cleavage site usage occurs at a time of profound
89 change in cell state in the male *Drosophila* germ line adult stem cell lineage. Spermatogenesis in
90 *Drosophila* starts with an oriented stem cell division that causes one of the daughter cells to be
91 displaced away from the somatic hub cells, stop self-renewing and start a program of transient
92 amplification by mitotic proliferation, founding a clone of interconnected spermatogonia
93 enclosed in two somatic cyst cells (Figure 1 B). After 4 rounds of mitosis the spermatogonia in
94 the cyst start differentiation, beginning with pre meiotic S phase, producing 16 interconnected
95 spermatocytes. After roughly 3.5 days of meiotic prophase, during which the germ cells grow 25

96 times in volume, the spermatocytes undergo meiosis I and II, producing a cyst of 64 haploid
97 spermatids. These then undertake a series of morphological changes that lead to production of
98 mature and functional sperm (Figure 1 B). The switch from spermatogonia to spermatocytes is
99 accompanied by changes in both cell state and gene expression. Almost all the transcripts
100 needed for the changes in cellular morphology that occur after meiosis are expressed in
101 spermatocytes (Raz *et al.* 2023). Changes in protein expression due to 3'UTR shortening by APA
102 may thus play key roles not only in setting up differences between proliferating spermatogonia
103 vs differentiating spermatocytes, but also in stage-specific translational repression or activation
104 required for proper timing of the subsequent cellular morphogenesis.

105 As the 3'UTR of mRNA molecules is a hub for *cis*-regulatory sequences that can attract
106 *trans*-acting factors such as miRNAs and RNA binding proteins to regulate mRNA stability and
107 translation efficiency, it is not surprising that the stage-specific 3'UTR shortening due to APA at
108 over 500 genes observed in spermatocytes compared to spermatogonia (Figure 1 C) resulted in
109 dramatic shifts in the proteins expressed between these two cell types. Polysome fractionation
110 revealed that transcripts from at least 200 genes switched from the long 3'UTR isoform co-
111 migrating with polysomes in spermatogonia to the short 3'UTR isoform not associated with
112 ribosomes in young spermatocytes. For another ~50 genes the long 3'UTR transcript isoforms
113 were not translated in spermatogonia while the short 3'UTR isoforms co-migrated with
114 polysomes in spermatocytes (Berry *et al.,* 2022). Thus, the developmental program(s) that
115 specify whether nascent transcripts are cleaved at proximal 3'UTR sites in spermatocytes rather
116 than the distal sites utilized in spermatogonia can switch expression of many proteins from ON
117 in spermatogonia to OFF in spermatocytes, or vice versa, from OFF in spermatogonia to ON in
118 spermatocytes.

119 Here we show that cell-type-specific production of transcripts with short 3'UTRs from
120 specific genes due to APA in spermatocytes is developmentally regulated by changing the levels
121 of one of the 4 complexes involved in 3' end cleavage and polyadenylation. We found that the
122 two protein components of CFII, encoded by the *PCF11* and *crowded by cid (cbc)* genes, are
123 upregulated in spermatocytes compared to spermatogonia. When levels of *PCF11* or *cbc* were
124 knocked down in spermatocytes by cell-type-specific RNAi, over 240 (~48%) of the ~500 genes

125 for which nascent transcripts would normally be cleaved at a proximal PAS in spermatocytes
126 instead resumed producing transcripts cleaved at the distal PAS utilized in spermatogonia.
127 Reciprocally, forced overexpression of CFII components in spermatogonia shifted 3' end
128 cleavage to the more proximal site for some of the transcripts from genes that undergo stage-
129 specific APA identified by Berry *et al.* (2022). Lowering protein levels in spermatocytes of
130 components of the cleavage machinery associated complexes CFI or CstF had much less effect
131 on the choice of 3' end cleavage site, suggesting that CFII plays a special role in regulating APA in
132 spermatocytes compared to spermatogonia. The PCF11 and Cbc proteins interact structurally:
133 immunoprecipitation of Cbc from testis extracts brought with it a >250 KDa form of PCF11.
134 Reduced expression of *PCF11* protein in spermatocytes due to RNAi resulted in concomitant
135 reduction in Cbc, suggesting that Cbc protein was not stable without its binding partner in CFII.
136 However, the reciprocal was not true. Plentiful PCF11 protein persisted in the nuclei of
137 spermatocytes in which expression of *cbc* had been knocked down by RNAi. Strikingly, *PCF11*
138 mRNA is broadly expressed in many cell types in adult flies, while *cbc* expression is low in most
139 adult tissues but upregulated in spermatocytes. These observations suggest a model where a
140 developmentally regulated increase in expression of *cbc* and *PCF11* in spermatocytes leads to
141 greater activity of cleavage factor CFII. This in turn facilitates recognition of an upstream PAS
142 and consequent 3'UTR shortening by APA of nascent transcripts from specific genes, indicating
143 CFII as a key player in developmental regulation of alternative cleavage and polyadenylation.

144 **Results**

145 **Specific cleavage factor proteins are upregulated in spermatocytes.**

146 Immunostaining of whole mount *Drosophila* testes revealed that some but not all
147 cleavage factor proteins were upregulated in spermatocytes (Figure 1 D-J). Relative expression
148 levels of six protein components of various sub-complexes of the cleavage factor machinery
149 were measured by immunostaining using available antibodies against endogenous proteins (for
150 CstF50, CstF64, PABP2 and PCF11), anti-HA (for *HA:cbc,* an available *Drosophila* line tagged with
151 HA at the endogenous *cbc* locus by CRISPR (Wu *et al.* 2021)), or anti-GFP (for an available
152 *Drosophila* line carrying a Fosmid based transgenic copy of CPSF6 tagged with sGFP (super
153 folder GFP) (Sarov *et al.* 2016)). In comparing nuclear signal in spermatogonia and
154 spermatocytes, each measurement was divided by the area of the corresponding nucleus to
155 account for the great difference in nuclear size between the two cell types.
156 Immunofluorescence signal from antibodies recognizing two members of the CstF complex,
157 CstF50 (Figure 1 E) and CstF64 (Figure 1 F), decreased slightly or remained almost the same in
158 nuclei in spermatocytes compared to nuclei in spermatogonia (quantified in Figure 1 D). In
159 contrast, immunofluorescence signal from antibodies recognizing PABP2 (Figure 1 G), the CFI
160 component CPSF6 (Figure 1 H) and the two members of CFII, Cbc (Figure 1 I) and PCF11 (Figure
161 1 J) increased substantially in nuclei from spermatocytes compared to nuclei from
162 spermatogonia (quantified in Figure 1 D).

163

164 **Reduction of CFII components levels in spermatocytes shifted APA to more distal sites.**

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

171 Immunofluorescence staining revealed that *bamGal4* driven knock down using the
172 *VDRC-103710* RNAi line lowered levels of *PCF11* protein detected in spermatocyte nuclei to a

173 level similar to that detected in spermatogonia (Figure 2 B). Testes from *UAS>PCF11-RNAi-VDRC-*
174 *103710;bamGal4* males viewed in live squash preparations by phase contrast microscopy had
175 plentiful spermatocytes but few elongating spermatid bundles (Figure 2 E). An RNAi line
176 directed against a different region of the PCF11 transcript, *UAS*>PCF11-RNAi-*NIG-10228R-4*, also
177 lowered the level of anti-PCF11 immunofluorescence signal in spermatocytes when driven using
178 *bamGal4* (Figure 2 C). Testes from *UAS*>PCF11-RNAi-*NIG-10228R-4;bamGal4* males viewed by
179 phase contrast microscopy had some elongating spermatid bundles (Figure 2 F). In both cases,
180 PCF11 protein was still detected in early spermatogonia and in nuclei of the somatic cyst cells
181 that accompany spermatocyte cysts, as expected since the *bamGal4* driver is not expressed in
182 these cells (Figure 2 B, C, arrow heads).

183 Knock down of *cbc*, the partner of PCF11 in CFII, also caused defects in spermatogenesis.
184 Immunofluorescence staining of testes from males heterozygous for a *cbc* allele tagged at the
185 endogenous locus with the HA epitope (*HA:cbc*) showed strong anti-HA signal in spermatocyte
186 nuclei, compared to little or no signal detected in spermatogonia. However, for two RNAi lines
187 directed against different regions of *cbc*, knock down under control of *bamGal4* reduced the
188 level of anti-HA signal in spermatocytes to below the level of detection, as in spermatogonia
189 (Figure 2 G-I). In both cases, analysis of unfixed squash preparations by phase contrast
190 microscopy showed that the knock down testes had abundant spermatocytes but lacked
191 elongating spermatid bundles (Figure 2 K, L).

192 *In situ* hybridization revealed that knock down of *PCF11* or *cbc* in spermatocytes by RNAi
193 affected APA at both the *Kpc2* (also named *isopeptidase-T-3*) and the *Red* genes (Figure 3). *Kpc2*
194 and *Red* had both been found to express transcripts with a long 3'UTR in testis filled with
195 spermatogonia but a much shorter 3'UTR in testes filled with mid-stage differentiating
196 spermatocytes (Figure 3 A, E, J). The relative levels of transcripts with long vs. short 3'UTRs
197 expressed in spermatogonia was compared by Hybridization Chain Reaction Fluorescence *in situ*
198 Hybridization (HCR-FISH), a technique that can be used to quantify relative RNA levels in a
199 sample, as the amplified fluorescent signal has been shown to be proportional to the number of
200 mRNA molecules in the quantified area (Trivedi *et al.* 2018; Choi *et al.* 2018; Schulte *et al.* 2024).
201 Comparing signal in the same sample from two different sets of HCR-FISH probes (SOM table 1),

202 one recognizing the common coding sequence and one recognizing the 3'UTR extension (Figure
203 3 A), revealed the proportion of the total transcripts from a given gene (recognized by the
204 common probe set) that contained the 3' UTR extension (recognized by the 3'UTR extension
205 probe). For both *Kpc2* and *Red*, knock down of either *PCF11* or *cbc* in spermatocytes under
206 control of *bamGal4* substantially increased the level of signal from the 3'UTR extension
207 detected in spermatocytes (Figure 3 C, D, H, I) compared to testes from control flies carrying the
208 *bamGal4* driver but not the RNAi transgene raised under the same temperature regimen (Figure
209 3 B, G). Quantification of fluorescent signal confirmed that for both *Kpc2* and *Red* the ratio of
210 fluorescence from probes recognizing the long 3'UTR versus probes against the coding
211 sequence (CDS) for each gene was significantly higher in the *PCF11* or *cbc* knock down
212 spermatocytes than in spermatocytes from control testes (Figure 3 F, K).

213 Analysis of cleavage site usage by 3'seq from testes in which either *PCF11* or *cbc* had
214 been knocked down in spermatocytes by RNAi under control of *bamGal4* showed that the
215 increase in signal from the 3'UTR extension observed in knock down spermatocytes was due to
216 a shift from cleavage at the proximal site normally utilized in spermatocytes to the more distal
217 cut site normally utilized in spermatogonia for both *Kpc2* and *Red* (Figure 4 A, B). Knock down of
218 *cbc* had the strongest effect. Knock down of *PCF11* shifted 3' end cleavage to the same distal
219 site as *cbc* knock down but to a lesser degree, especially with the *UAS>PCF11-RNAi-NIG-*
220 *10228R-4* construct (Figure 3 A, B), consistent with the less efficient reduction in PCF11 protein
221 levels observed with that particular line (Figure 2 C).

222 3'seq data revealed that loss of function of *cbc* or *PCF11* affected APA at many loci. For
223 roughly half of the ~500 genes previously identified by Berry *et al.* (2022) as undergoing 3'UTR
224 shortening by APA in spermatocytes compared to spermatogonia, knock down of either *PCF11*
225 or *cbc* in spermatocytes resulted in increased usage of the distal cleavage site normally used in
226 spermatogonia (Figure 4 C). When 3'seq data was analyzed via PolyA miner (Yalamanchili *et al.*
227 2020) providing as input the list of pairs of proximal PAS (pPAS) and distal PAS (dPAS) from the
228 >500 genes undergoing APA identified by Berry *et al.* (2022), many genes showed increased
229 usage of the distal PAS in knock down testes compared to control testes. Orange in Figure 4C
230 indicates that the frequency of nascent transcripts from a given gene being cleaved at the distal

231 PAS in knock down testes sample was higher than in control testes (positive PolyA index), while
232 blue indicates that the frequency of the nascent transcripts being cleaved at the proximal PAS in
233 knock down testes was higher than in control testes (negative PolyA index). Noticeably, a
234 negative PolyA index does not indicate a shortening event but indicates a gene for which the
235 proximal PAS was overrepresented in knock down compared to control samples. This could be
236 caused either by different expression of the gene compared to in control testes or by the
237 different cell composition in wild-type vs. knock down testes, as generally knock down testes do
238 not have as many elongating spermatids as control testes. Again, knock down of *cbc* in
239 spermatocytes had the strongest effect on the APA switch, affecting APA at the highest number
240 of genes. Importantly, >90% of the genes where APA was affected by knock down of *PCF11*
241 (PolyA index > 0.5, indicating more cleavage at the distal PAS) also showed a switch in APA to
242 more cleavage at the distal site with knock down of *cbc,* consistent with these two CFII
243 components acting together in the choice of where to make the 3' end cut for transcripts from
244 specific genes. Analysis of RNAseq libraries from testes samples in which *cbc* or *PCF11* were
245 knocked down under *bamGal4* revealed that in general the genes that had a PolyA index either
246 greater than 0.5 or smaller than negative 0.5 were not expressed at dramatically different levels
247 in knock down versus control testes (SOM Figure 3).

248 3'seq revealed that lowering the levels of the nuclear PABP2 protein in spermatocytes
249 by RNAi also frequently affected 3'UTR shortening by APA (Figure 4 A-C). Immunofluorescence
250 staining showed effective knock down of PABP2 protein to below detection in spermatocytes
251 from flies carrying *UAS>PABP2-RNAi-VDRC-106466;bamGal4*, although PABP2 protein was still
252 abundant in nuclei of somatic cyst cells, as expected (SOM Figure 4 B). For two different RNAi
253 lines, testes from *UAS>PABP2-RNAi;bamGal4* flies had plentiful spermatocytes but lacked
254 elongating spermatid bundles (SOM Figure 4 D, E). 3'seq revealed that in testes from
255 *UAS>PABP2-RNAi-VDRC-106466;bamGal4*, a substantial percentage of transcripts from the *Kpc2*
256 and *Red* loci switched 3' end cleavage from the proximal site normally utilized in spermatocytes
257 to the more distal site normally utilized in spermatogonia (Figure 4 A, B). Analysis of the 3'seq
258 data showed that for the >500 genes identified previously as undergoing APA in spermatocytes
259 compared to spermatogonia,~22% showed increased usage of the distal PAS when PABP2 was

260 knocked down in spermatocytes. Furthermore, most (~92%) of the genes that switched toward
261 the more distal PAS (PolyA index > 0.5) after knocking down *PABP2*, also showed a shift (PolyA
262 index > 0.5) towards the distal PAS when either *PCF11* or *cbc* was knocked down in
263 spermatocytes (Figure 4 C, F).

264 In contrast, knock down of any of the three components of the CstF complex (*CstF64*,
265 *CstF50*, and *su(f)*) or a component of CFI (*CPSF6*) in spermatocytes had much less effect on the
266 APA profile of mRNAs from the ~500 genes previously shown to undergo 3' end shortening due
267 to APA in spermatocytes, as assessed by 3'seq (Figure 4 C). Immunofluorescence staining with
268 antibodies against CstF64 or CstF50 showed that levels of the proteins were substantially
269 reduced in spermatocytes in testes in which expression of the respective genes had been
270 knocked down by RNAi under control of *bamGal4* (SOM Figure 5). In both cases, signal
271 remained in the nuclei of early spermatogonia and cyst cells, as expected from the cell-type-
272 specificity of the *bamGal4* expression driver. No antibody reagents against *su(f)* were available
273 to assess effectiveness of the RNAi construct directed against *su(f)*. However, the defects in
274 spermatogenesis in *UAS>su(f)-RNAi;bamGal4* testes observed by phase contrast microscopy
275 (SOM Figure 5) and the effects of knocking down *su(f)* on protein levels of CstF50 in
276 spermatocytes nuclei (SOM Figure 6), suggested that the knock down of *su(f)* was effective.
277 Levels of nuclear CstF50 protein drastically decreased in spermatocytes from males carrying an
278 RNAi construct directed against *su(f)* expressed under control of *bamGal4* (SOM Figure 6 D).
279 Knock down of *CstF64* under control of *bamGal4* also lowered the levels of CstF50 in
280 spermatocyte nuclei (SOM Figure 6 F). Notably, when either *su(f)* or *CstF64* were knocked down
281 in spermatocytes, nuclear CstF50 staining decreased, although some signal remained in the
282 cytoplasm. Together the protein expression data suggest that components of the CstF complex
283 may depend on each other for protein stability and/or nuclear localization, as previously
284 reported in HeLa cells (Grozdanov *et al.* 2018). Consistent with the knock downs affecting
285 function, squash preparations of testes from *UAS>su(f)-RNAI-BL-65693;bamGal4* showed
286 abundant spermatocytes but no elongating spermatid stages (SOM Figure 5 G). Testes from
287 *UAS>CstF64-RNAi-BL65987;bamGal4* males also showed defective spermatogenesis, with
288 plentiful spermatocytes but extremely aberrant elongation stages (SOM Figure 5 F).

289 Surprisingly, even though CstF50 protein was clearly strongly reduced in spermatocytes from
290 *UAS>CstF50-RNAi-BL77377;bamGal4* males, spermatogenesis appeared largely normal by phase
291 contrast microscopy (SOM Figure 5 C).

292 Knock down of either *CstF64* or *su(f)* in spermatocytes shifted the site of 3' end
293 formation toward the more distal site utilized in spermatogonia at some of the >500 genes
294 previously identified as subject to 3'UTR shortening in spermatocytes due to developmentally
295 regulated APA (Figure 4 C). These genes tended to be among the same set of genes where APA
296 was affected by knock down of components of CFII: ~95% of the genes where APA was affected
297 by knock down of *CstF64* were also affected by knock down of either *PCF11, cbc,* or both (Figure
298 4 G). However, the effects of loss of function of CstF components on APA in spermatocytes was
299 small compared to the effect of loss of CFII components: only 44 of the 258 genes where APA
300 was strongly affected by knock down of either *cbc* or *PCF11* in spermatocytes also showed a
301 switch toward the more distal cut site in testes where function of *CstF64* had been knocked
302 down in spermatocytes (Figure 4 G). More than 85% of the genes affected (PolyA index > 0.5) by
303 *CstF64* knock down were also affected either by *CstF50* or *su(f)* knock down, and 68% of the
304 genes affected by *su(f)* knock down were also affected by either *CstF64* or *CstF50* knock down
305 (Figure 4 H), consistent with these three components acting together. For *CPSF6*, a component
306 of the CFI complex, which binds nascent transcripts upstream the PAS, knock down under the
307 control of *bamGal4* caused extensive defects in spermatid differentiation for one RNAi line
308 (VDRC 107147) and variable or weaker defects for the other RNAi line tested (BL 34804, used for
309 3'seq), suggesting partial loss of function (SOM Figure 7). Analysis of cut sites utilized by 3' end
310 seq indicated that for the >500 genes previously identified as undergoing APA in spermatocytes,
311 the effect of *CPSF6* knock down on APA using the BL 34804 RNAi line was weak compared to
312 knock downs of *cbc* or *PCF11* (Figure 4 C). However, the weak effect could in part be due to a
313 weaker knock down.

314 *De novo* genome wide calling of PAS sites located on annotated 3'UTRs revealed that
315 knock down of components of the cleavage machinery affected APA at additional loci (SOM
316 Figure 8 and SOM Figure 9) in addition to the >500 genes from Berry *et al.* (2022). For this
317 analysis we only considered genes in which the most upregulated PAS and the most

318 downregulated PAS in a knock down compared to a wild-type sample were located in the
319 annotated gene 3'UTR. Again, the strongest effects were detected in *cbc* knock down testes,
320 followed by *PCF11* knock down and *PABP2* knock down testes (SOM Figure 8), while *CstF64*,
321 *CstF50*, *su(f)* and *CPSF6* knock downs had much milder effects on the global APA profile (SOM
322 Figure 9).

323 Of the components of the cleavage machinery tested, the 3'seq data indicated that the
324 3'UTR shortening of transcripts from selected genes due to APA in spermatocytes was most
325 strongly dependent on levels of CFII (composed of Cbc and PCF11), or *PABP2*. For almost half of
326 the >500 genes previously identified as undergoing the developmentally regulated APA,
327 lowering function of one or the other of these proteins in spermatocytes resulted in a shift
328 toward usage of the more distal 3' end cut site typical of spermatogonia. Several genes seemed
329 to be more affected by knock down of *cbc* than by knock down of *PCF11*, for example *oaf*
330 (Figure 4 D). Notably, over half of the genes previously identified as subject to 3'UTR shortening
331 in spermatocytes compared to spermatogonia due to APA were not substantially affected by
332 knock down of any of the cleavage factor machinery we tested, see *nudE* for example (Figure 4 E
333 and SOM Figure 2).

334

335 **Transcription of the CFII components *PCF11* and *cbc* are strongly upregulated in**
336 **spermatocytes.**

337 According to Fly Base, the *PCF11* locus can give rise to 7 different RNA isoforms (Figure 5
338 A), grouped in 3 classes for simplicity: a short isoform (*PCF11-RC*) of 819 nucleotides predicted
339 to encode a small protein (~10 KDa), a medium isoform (*PCF11-RB*) of 2647 nucleotides
340 encoding a ~63 KDa predicted protein, and a group of ~6000 nucleotide transcripts (*PCF11-RD-*
341 *RE-RF-RG-RH*) predicted to encode a family of >200 KDa proteins that have subtle differences in
342 a central region. All the transcript isoforms share the same 5'UTR but differ in their 3'UTR (as
343 well as in exon combinations). PCF11 has a CID domain (POLII C-terminal binding domain)
344 present in both the PCF11-long and the 63 KDa PCF11-RB isoforms, but only partially present in
345 the short protein encoded by *PCF11-RC*. PCF11 homologs in mammals and yeast have been
346 shown to bind Cbc homologs through a region close to the C-terminus of PCF11 (Sadowski *et al.*

347 2003; Schäfer *et al.* 2018). Notably, while the *PCF11-RNAi-NIG10228-R4* line targets a region
348 common to all *PCF11* mRNA isoforms, *PCF11-RNAi-VDRC107310* targets a region present only in
349 the *PCF11-long* isoforms. The *cbc* locus is simpler, with only one predicted transcript 1654
350 nucleotides long, encoding a 46.8 KDa predicted protein (Figure 5 A).

351 The two *Drosophila* components of CFII show substantially different patterns of
352 transcript expression in adult flies. In the Fly Cell Atlas single nuclear RNA-seq analysis of
353 tissues from adult flies (Li *et al.,* 2022), transcription of *PCF11* was detected at generally high
354 levels in many cell types (Figure 5 B). In contrast, *cbc* expression was detected at substantial
355 levels in only a small subset of nuclei from adult tissues, with the clusters showing the highest
356 *cbc* transcription belonging to the testes plus seminal vesicle sample (Figure 5 C, D). Plotting
357 expression on the UMAP of snRNA-Seq data from testis plus seminal vesicle revealed that *PCF11*
358 is transcribed in many somatic cell types in the testis, including terminal epithelia, seminal
359 vesicle, nuclei from testis muscle cells and prominently in the somatic cyst cells that enclose
360 germ line cysts. *PCF11* transcripts were also expressed in germ cells, with the level increasing
361 substantially in spermatocytes compared to spermatogonia (Figure 5 E). Transcription of *cbc*
362 also increased substantially in spermatocytes compared to spermatogonia (Figure 5 F).

363 Analysis of RNA expression in wild-type testes by HCR-FISH revealed that the long
364 transcripts from *PCF11* were upregulated in spermatocytes compared to spermatogonia (Fig. 5
365 H-H'-H''). In contrast, the intermediate sized transcript isoform *PCF11-RB*, detected with a set
366 of probes recognizing the 3'UTR of *PCF11-RB*, which derives from an intronic region of the
367 *PCF11-long* isoforms so should not recognize fully processed *PCF11-long* isoforms in the
368 cytoplasm, was extremely low in both spermatogonia and spermatocytes (Figure 5 I-I''). Single
369 nuclear RNAseq from the Fly cell Atlas confirmed that while the collective *PCF11-long* isoforms
370 are highly expressed in spermatocytes, cyst cells and somatic cell nuclei in the testis (Figure 5 K),
371 the *PCF11-RB* transcript was only lowly expressed in the germline, with higher but still modest
372 expression in cyst cell and other somatic cell nuclei in the testis (Figure 5 L). Similar HCR-FISH
373 using a set of probes against the *cbc* protein coding region confirmed that expression of *cbc*
374 mRNA was strongly upregulated in spermatocytes compared to spermatogonia (Figure 5 J-J'').
375

376 **PCF11 and Cbc co-immunoprecipitate and knock down of *PCF11* causes decrease of Cbc**
377 **protein in spermatocyte nuclei.**
378 Analysis of proteins expressed in wild-type testes by polyacrylamide gel electrophoresis
379 (PAGE) followed by Western blot probed with antibodies against PCF11 showed a predominant
380 high molecular weight (>250 KDa) band (Figure 6 A), consistent with the expression of *PCF11-*
381 *long* mRNA isoforms detected by HCR-FISH and snRNA-seq (Figure 5 H, K). Anti-PCF11 also
382 detected a ~70 KDa protein isoform in testis extracts, as predicted from the *PCF11-RB* mRNA
383 isoform. Although the antibody used was raised against amino acids 1-283 of *Drosophila* PCF11
384 (Zhang and Gilmour 2006) no protein product corresponding to PCF11-RC was observed,
385 despite a seemingly high level of the *PCF11-RC* transcript isoform detected by snRNAseq of
386 *Drosophila* testis tissue (SOM Figure 10). Western Blots of testes from flies homozygous for the
387 *HA:cbc* allele probed with anti-HA showed expression of the expected ~50 KDa isoform (Figure 6
388 B).
389 Immunoprecipitation with anti-HA antibodies from extracts of testes from flies
390 homozygous for *HA:cbc* co-immunoprecipitated the >250 KDa band from the PCF11 isoforms,
391 consistent with the Cbc and PCF11 proteins interacting in a complex (Figure 6 C). The 70 KDa
392 isoform of PCF11 was not detected in the immunoprecipitate (entire gels of Western Blots and
393 co-IP experiments are reported in SOM 11). Consistent with physical interaction between the
394 Cbc and PCF11 proteins, accumulation of Cbc protein in spermatocytes was at least partially
395 sensitive to the levels of PCF11 protein. Analysis by RNA-seq showed that when PCF11
396 expression was knocked down in late spermatogonia and spermatocytes by RNAi under control
397 of *bamGal4*, *cbc* transcript was still expressed in *PCF11* knock down testes at a similar level
398 compared to control testes (Figure 6 D). Likewise, in testes in which *cbc* had been knocked down
399 by RNAi under control of *bamGal4*, transcripts from the *PCF11* locus were still expressed (Figure
400 6 E). However, the situation was different at the level of protein accumulation in spermatocytes.
401 Simultaneous immunostaining of testis from *HA:cbc/+;bamGal4* males with anti-PCF11 and anti-
402 HA antibodies showed both proteins present in spermatocyte nuclei from control males (Figure
403 6 F). However, for two different RNAi lines directed against *PCF11*, knock down of *PCF11* under
404 control of *bamGal4* resulted in drastic reduction of signal from HA:Cbc (Figure 6 G'-H',

405 quantified in K). Similar reduction of Cbc signal in *UAS>PCF11_VDRC-RNAi;bamGal4* and
406 *UAS>PCF11_NIG-RNAi;bamGal4* spermatocytes was observed in flies homozygous for the
407 *HA:cbc* tagged allele (SOM 12). In contrast, knock down of *cbc* by RNAi under control of
408 *bamGal4* did not cause loss of signal from anti-PCF11 (Figure 6 I''-J'', quantified in L).
409
410
411 **Forcing high expression of *cbc* and/or *PCF11* in spermatogonia induced cleavage at proximal**
412 **PASs normally used in spermatocytes**
413 Our data from the knock down experiments in spermatocytes suggest that the high
414 levels of CFII components expressed in spermatocytes are required for cleavage of nascent
415 transcripts of certain genes at an upstream PAS rather than at the distal PAS used in
416 spermatogonia. Forced overexpression of CFII components in spermatogonia triggered a
417 fraction of the nascent transcripts from certain genes to be cleaved at the upstream site
418 normally used in spermatocytes.
419 Minigenes with the coding sequences of *cbc* with a V5 tag or *PCF11-RF* with an HA-FLAG
420 tag, both under the control of two repeats of 5X UAS and the *hsp70* promoter, were introduced
421 into the *Drosophila* genome. These transgenes allow spermatogonial specific overexpression of
422 *cbc-V5* and/or *PCF11-RF-HA* by crossing males carrying the designed UAS constructs with
423 females carrying a combination of *nanosGal4* (to drive in stem cells and early spermatogonia)
424 and *bamGal4* (to drive in mid-late spermatogonia). Transgenes and Gal4 expressing lines were
425 introduced into a $bam^{-/-}$ genetic background, in which early stage spermatogonia over
426 proliferate and fail to differentiate into spermatocytes and expression from the transgenes
427 scored by immunofluorescence staining with anti-V5 and anti-HA (Figure 7 A-D). Anti-HA
428 staining to visualize the overexpressed PCF11-RF-HA showed that the protein was highly nuclear
429 in spermatogonia from flies overexpressing *PCF11-HA-RF* (Figure 7 B'') and from flies
430 overexpressing simultaneously *cbc-V5* and *PCF11-RF-HA* (Figure 7 D''), recapitulating PCF11
431 nuclear localization in wild-type spermatogonia (Fig 1 J). Anti-V5 staining to visualize the
432 overexpressed Cbc-V5 was instead both nuclear and cytoplasmic in spermatogonia from
433 transgenic flies overexpressing *cbc-V5* (Figure 7 C'). However, in spermatogonia overexpressing

434 both *PCF11-RF-HA* and *cbc-V5*, more of the anti-V5 signal was nuclear and less cytoplasmic
435 (Figure 7 D'). Quantification of the fluorescent signal from the nucleus and cytoplasm showed
436 that the ratio of anti-HA signal in the nucleus compared to the cytoplasm remained constant
437 when *PCF11-RF-HA* was overexpressed alone or with *cbc-V5*. However, the ratio of nuclear to
438 cytoplasmic anti-V5 signal increased when *cbc-V5* was overexpressed with *PCF11-RF-HA*
439 compared to when *cbc-V5* was overexpressed alone (Figure 7 E). Immunoprecipitation of Cbc-
440 V5 using anti-V5 antibody from extracts of *bam*$^{-/-}$ testes in which *PCF11-RF-HA* and *cbc-V5* were
441 both overexpressed in spermatogonia under control of *nanosGal4* and *bamGal4* brought down
442 PCF11-RF-HA, indicating that the overexpressed tagged proteins interact, similar to their
443 counterparts expressed in wild-type testes (Figure 7 F, entire gels in SOM Figure 13).
444 Analysis of 3'seq libraries from *bam*$^{-/-}$ testes in which *cbc-V5* alone, *PCF11-RF-HA* alone,
445 or *cbc-V5* and *PCF11-RF-HA* together were overexpressed in spermatogonia compared with
446 3'seq libraries from *bam*$^{-/-}$ testes in which neither was overexpressed revealed that of the ~500
447 genes identified by Berry *et al.* (2022) as producing mRNAs with different 3'UTR length due to
448 APA, ~140 showed upregulation of 3' end cleavage at the proximal PAS (see for example the
449 genes *Rop, orb, Mitofilin* and *bor* (Figure 7 H-K). 47% of the genes that showed an increase in
450 usage of the proximal PAS normally used in spermatocytes (polyA index < -0.5) in samples in
451 which either *cbc-V5* or *PCF11-RF-HA* or both were overexpressed in *bam*$^{-/-}$ spermatogonia also
452 showed higher usage of the distal PAS in samples in which *cbc* or *PCF11* had been knocked
453 down in spermatocytes (Figure 7 G). The increased processing at proximal PASs in *bam*$^{-/-}$ testes
454 in which CFII components were overexpressed only affected some of the nascent transcripts
455 from a given locus and was never an APA switch as striking as the one observed when *cbc* or
456 *PCF11* were knocked down in spermatocytes (SOM Figure 14 and Figures 3 and 4).

457 **Discussion**

458 Recent results show that stage-specific Alternative Cleavage and Polyadenylation (APA)
459 can regulate dramatic changes in the suite of proteins expressed as cells advance stepwise
460 through a differentiation sequence (Berry *et al.* 2022). The developmental mechanisms that
461 regulate cell type and stage-specific APA can thus be key contributors to cellular differentiation.
462 Our analysis has revealed that cell-type-specific differential expression of components of
463 Cleavage Factor complex II (CFII) of the 3' end cleavage machinery regulates alternative
464 polyadenylation in the *Drosophila* male germ line adult stem cell lineage. Results of cell-type-
465 specific knock down experiments suggest that the developmentally specified increase in
466 expression of CFII components PCF11 and Cbc is important for the shift to 3' end cleavage at a
467 more proximal site in spermatocytes compared to spermatogonia at over 200 loci during
468 *Drosophila* spermatogenesis. Forcing overexpression of CFII components in spermatogonia
469 resulted in increased 3' end cleavage at the proximal site more typical of spermatocytes for a
470 fraction of transcripts from some genes. Knock down of several other cleavage factors in
471 spermatocytes had much milder effects on shifting the 3' end cleavage from the proximal site to
472 the more distal site normally utilized in spermatogonia, singling out CFII for a key role in the
473 developmentally regulated 3'UTR shortening due to APA characteristic of wild-type *Drosophila*
474 spermatocytes.

475 Cell-type-specific upregulation of *cbc* may be a key factor in developmental regulation of
476 APA in differentiating male germ cells. While *PCF11* transcripts are broadly expressed in fly adult
477 tissues, *cbc* seems to be particularly highly expressed mostly in spermatocytes. In addition,
478 previous studies involving knock down of *PCF11* in cancer cells and zebrafish embryos showed
479 that downregulating *PCF11* broadly induced 3'UTR lengthening. On the other hand, knocking
480 down *CLP1* (the mammalian homolog of *cbc*) in mammalian cells did not show remarkable
481 effects on APA (Li *et al.* 2015; Ogorodnikov *et al.* 2018; Kamieniarz-Gdula *et al.* 2019). However,
482 it is possible that *cbc* was not highly expressed or did not play a key role in regulating APA in the
483 cell types tested. In contrast, knock down of *cbc* in *Drosophila* spermatocytes strongly affected
484 the APA profile, resulting in cleavage at the distal site typical of spermatogonia for almost half of
485 the genes that would instead normally be cleaved at proximal sites in spermatocytes.

486 Several lines of evidence suggest that PCF11 may carry out functions independently of
487 CFII. PCF11 has been found to bind the C-Terminal Domain (CTD) of RNA POLII *in vitro* and in
488 yeast (Sadowski *et al.* 2003; Zhang and Gilmour 2006), where it acts as a termination factor
489 (Guéguéniat *et al.* 2017; Schäfer *et al.* 2018; Kamieniarz-Gdula *et al.* 2019). ChiP-seq of POLII in
490 control versus *PCF11* knock down human cells showed that POLII fails to terminate efficiently
491 when PCF11 is reduced (Kamieniarz-Gdula *et al.* 2019). Accordingly, downregulating *PCF11*
492 expression levels has been shown to induce 3'UTR lengthening in cancer cells and zebrafish
493 embryos (Li *et al.* 2015; Ogorodnikov *et al.* 2018; Kamieniarz-Gdula *et al.* 2019). Whether
494 skipping upstream cleavage sites might be linked with the dynamics of POLII termination, and
495 what other functions PCF11 might carry out in association with RNA POLII still need to be
496 elucidated. PCF11 has been shown to bind RNA *in vitro*, with a preference for G-rich regions. On
497 the other hand, Cbc does not have a recognized RNA binding domain, and immunoprecipitation
498 of Cbc *in vitro* failed to detect bound RNA (Schäfer *et al.* 2018). Our findings raise the possibility
499 that Cbc might act as a developmentally regulated factor that, when upregulated in certain cell
500 types, might bind to PCF11 to form CFII, utilize the ability of PCF11 to recognize RNA, and
501 modulate PCF11 activity to enhance 3' end cleavage at specific sites.
502 Notably, not all of the ~500 genes found to undergo cell-type-specific 3'UTR shortening
503 due to APA in *Drosophila* spermatocytes were affected by *PCF11* or *cbc* knock down, suggesting
504 that other developmental mechanisms (such as changes in transcription rates and/or activity of
505 other RNA binding proteins) might play a role in regulating stage-specific APA at some genes. In
506 addition, forced overexpression of CFII components in spermatogonia increased 3' end cleavage
507 at the proximal PAS normally used in spermatocytes only for some genes, suggesting that
508 additional factors (perhaps PABP2, for example) may help drive cleavage at the proximal PAS. It
509 is also possible that CFII has spermatocyte specific components or interacting proteins that still
510 need to be identified.
511 PCF11 has been found to be part of an autoregulatory feedback loop in cancer cells and
512 zebrafish, where it auto-regulates its own protein levels by activating a premature PAS
513 (Kamieniarz-Gdula *et al.* 2019). Our 3' end sequencing data revealed that a PAS located very far
514 5' in the *PCF11-RC* coding sequence becomes highly utilized in spermatocytes compared to

515 spermatogonia. Interestingly, cleavage at this premature PAS was dependent on *PCF11* and *cbc*
516 levels, as the upstream PAS was no longer highly utilized if *PCF11* or *cbc* were knocked down in
517 spermatocytes (SOM Figure 15). This could mimic what has been observed in human cells and
518 zebrafish embryos and suggests that the proposed PCF11 auto-regulatory loop might be
519 dependent on both components of the CFII complex rather than on PCF11 alone.

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

530 **Materials and Methods**

531

532 **Fly Strains, husbandry and design of UAS transgenes for overexpression.**

533 Flies were grown on standard molasses media and kept at 25 °C, 29 °C or room temperature as
534 indicated. For RNAi induced knock downs, fly crosses (*bamGal4* driver females X RNAi line
535 males) were kept at 25 °C for three days. Adults were then shifted to a new media bottle and
536 kept at 25 °C while the eggs and early larvae were shifted to 29 °C to increase knock down
537 efficiency. To further increase knock down penetrance, we used *Gal4* drivers that also expressed
538 *UAS > dicer2*. For over expression in spermatogonia of *PCF11* and *cbc*, male flies expressing
539 minigenes of *PCF11-RF*-HA-FLAG and/or *cbc*-V5 under the control of two 5XUAS sequences and
540 a *hsp70* promoter were crossed with females bearing both the *nanos*Gal4 and *bam*Gal4
541 transgenes to induce expression in both early (*nanosGal4*) and late (*bamGal4*) spermatogonia.
542 Crosses were kept three days at 25 °C. Then the adults were removed, and the eggs and larvae
543 shifted to 29 °C.

544 Plasmid manipulation was performed following standard cloning strategies. In brief, for
545 the plasmid to force expression of *PCF11-RF-HA*, the plasmid LD11480 from DGRC (Drosophila
546 Genomics Resource Center) was used as a template, while for *cbc-V5* forced expression the
547 plasmid UFO01678 from DGRC was used as a template. Plasmid injection and final line
548 balancing were performed by BestGene. The construct to overexpress *PCF11-RF-HA* was
549 inserted on an attP landing site on the left arm of the second chromosome, while the construct
550 to overexpress *cbc-V5* was inserted on an attP landing site on the right arm of the second
551 chromosome. Plasmid maps are reported in supplementary material. Fly lines used in this study:
552 *W1118* (BL #3605), *bamGal4* (Chen and McKearin 2003), *nanosGal4-VP16* (Doren *et al.* 1998),
553 bam^{1}/TM6B (McKearin and Spradling 1990), bam^{86}/TM6B (McKearin and Ohlstein 1995),
554 *CPSF6:GFP* (VDRC #318105), *HA:Cbc* (N-terminal CRISPR tag, kindly gifted by the Wang lab, (Wu
555 *et al.* 2021)), *PCF11*-RNAi_1 (VDRC #103710) *PCF11*_RNAi_2 (NIG #10228R4), *cbc*-RNAi_1 (BL
556 #77426), *cbc*-RNAi_2 (VDRC #100686), *CPSF6*-RNAi_1 (BL #34804), *CPSF6*-RNAi_2 (VDRC
557 #107147), *CstF64*-RNAi_1 (BL #65987), *CstF50*-RNAi_1 (BL #77377), *PABP2*-RNAi_1 (VDRC

558 #106466), *PABP2*-RNAi_2 (VDRC #33499), *su(f)*-RNAi_1 (BL 55693), UAS > *PCF11-RF*-HA-FLAG
559 (this study), UAS > *cbc*-V5 (this study).
560
561 **Immunofluorescence (IF) staining and signal quantification.**
562 Testes from 0-3 days old males were removed by dissecting in 1X PBS on a cyclops dissecting
563 dish at room temperature over 30 minutes or less, then testes were immediately fixed in a 1.7
564 mL Eppendorf tube with 1 mL 5% formaldehyde in 1X PBS for 30 minutes at room temperature
565 with rocking. If needed, fixed samples were stored in 1X PBS at 4 ^{o}C for up to one week. Testes
566 were then permeabilized in 0.3% Triton-X-100 and 0.6% sodium deoxycholate in 1X PBS for 30
567 minutes at room temperature, rinsed twice in 1X PBS + 0.05% Tween-20 (PBT), blocked for 30
568 minutes in 1X Western Blocking Reagent (WBR, Roche #11921673001) in PBT at room
569 temperature, then incubated with primary antibody mixed in 1X WBR+PBT overnight at 4 ^{o}C.
570 After primary antibody incubation, testes were rinsed twice in PBT, blocked again for 30 minutes
571 in 1X WBR+PBT at room temperature, then incubated with secondary antibody mixed in 1X
572 WBR+PBT for 2 hours at room temperature in a dark container. Testes were then rinsed twice in
573 PBT and mounted on a glass slide in 10 uL of ProLong diamond anti-fade with DAPI
574 (ThermoFisher, CAT #P36962). Primary antibodies used: guinea pig anti-CstF50 (1:300, kindly
575 gifted by John T. Lis (Ni *et al.* 2008)), rabbit anti-PABP2 (1:500, kindly gifted by Martine
576 Simonelig (Benoit *et al.* 1999)), goat anti-Vasa (1:200, Santa Cruz Biotechnology sc-26877,
577 discontinued), rat anti-Vasa (1:10 for supernatant or 1:100 for the concentrated version, both
578 from DSHB, antibody ID 760351) rabbit anti-CstF64 (1:400, kindly gifted by Zbigniew Dominski
579 (Skrajna *et al.* 2018)), rabbit anti-PCF11 (1:200, kindly gifted by David Scott Gilmour (Zhang and
580 Gilmour 2006)), rabbit anti-HA (1:500, Cell Signalling Technologies, #3724), mouse anti-HA
581 (1:200, Biolegends #901501), chicken anti-GFP (1:10.000, Abcam #13970) and mouse anti-V5
582 (1:200, Invitrogen #46-0705). Secondary antibodies used: donkey or goat anti-rabbit, anti-
583 mouse, anti-guineapig, anti-chicken, anti-rat and anti-goat, all used at 1:500 and conjugated
584 with either Alexa 488, Alexa 568, Alexa 594 or Alexa 647 depending on the combination of
585 antibodies used in each experiment (all from the ThermoFisher catalog). Imaging of IF
586 experiments was performed on a Leica SP8 Confocal microscope. Protein levels in Figures 1 and

587 6 were quantified in Fiji by manually delineating the cell nuclei and applying the formula:
588 *Integrated_Density – (Area X Background_Mean)*. The Nuclear-Cytoplasmic ratio of Cbc-V5 and
589 PCF11-RF-HA proteins in Figure 7 were quantified in Fiji by measuring the mean intensity of
590 protein signal in 10 different uniform pixels areas per image in the central region of the sample
591 and applying the formula: $(Mean_protein - Mean_background)_{Nucleus} / (Mean_protein -$
592 $Mean_background)_{Cytoplasm}$.
593
594 **HCR™-RNA-FISH.**
595 Testes were dissected from 0-3 day old males in 1X PBS on a cyclops dissecting dish at room
596 temperature over a maximum of 30 minutes, then fixed and permeabilized as done for
597 immunofluorescence. The protocol for "generic sample in solution" from the Molecular
598 Instruments website (https://www.molecularinstruments.com/hcr-rnafish-protocols) was
599 followed. Hybridization buffer, wash buffer, amplification buffer and hairpins were all from
600 Molecular Instruments (HCR™ RNA-FISH Bundle). After permeabilization as described for
601 immunofluorescence above, samples were incubated with 200 uL of probe hybridization buffer
602 for 30 minutes at 37 °C. Probe mixes were prepared by mixing 3.2 uL of 1 uM gene-specific-
603 probes mix (one probe mix per gene or transcript) in 200 uL of probe hybridization buffer.
604 Hybridization buffer was removed, probe mix added, and the samples were incubated overnight
605 in probe mix at 37 °C in a water bath. The following day samples were washed 4 times for 15
606 minutes each in 200 uL of probe wash buffer at 37 °C and then washed 3 times 5 minutes each
607 in 5X SSCT [5X sodium chloride sodium citrate (SSC, Invitrogen #AM9770) and 0.1% Tween-20 in
608 H_2O] at room temperature. Samples were then incubated in 500 uL of amplification buffer for
609 30 minutes at room temperature. In the meantime hairpin mix was prepared: 30 pmol per
610 sample (5 uL) of harpin h1 and hairpin h2 (each fluorophore has its own harpin combination, in
611 our case since we always imaged 2 transcripts at the same time, we used 2 fluorophores and
612 therefore 4 hairpins) were heated at 95 °C for 90 seconds in a thermocycler in the dark and
613 cooled down at room temperature in a drawer for 30 minutes (each hairpin was heated up and
614 cooled in a separate tube). To prepare hairpin mixes 5 uL of each harpin were added to 250 uL
615 of amplification buffer. The samples were then incubated in harpin mix at room temperature in

616 the dark overnight in a drawer with no rocking. The following day samples were washed with
617 500 uL 5X SSCT 2 times for 5 minutes each, then 2 times 30 minutes each and finally 1 time 5
618 minutes, always in the dark and at room temperature. Samples were then mounted on a glass
619 slide in 10 uL of ProLong diamond anti-fade with DAPI (ThermoFisher, CAT #P36962). Probes
620 were designed as in Bedbrook *et al.* 2023 and ordered from Integrated DNA technologies as
621 Oligo Pools (oPools™ 50 pmol/oligo). Harpins H1 and H2 were ordered conjugated with either
622 488 or 546 fluorophores from Molecular Instruments. Probe sequences are listed in
623 supplementary material. Imaging was performed on a Leica SP8 confocal microscope. For each
624 testis, 5 areas of spermatocyte cytoplasm from different testis regions were quantified on Fiji
625 from both probe channels, the background was subtracted and then the ratio between the
626 probe recognizing the 3'UTR extensions and the probe recognizing the protein coding region
627 was calculated. The following formula was used: *(Area_Mean_p2 – Back_Mean_p2) /*
628 *(Area_Mean_p1 – Back_Mean_p1)* where p2 is the probe set recognizing the 3'UTR extension
629 while p1 is the probe set recognizing the protein coding region of the transcript. Area's size
630 varied between pictures but was constant between the 5 measurements from the same image.
631
632 **3' end sequencing and analysis.**
633 Testes were dissected from 0-3 day old males in S2 media in batches of 20-80 flies in a cyclops
634 dissecting dish at room temperature over a maximum of 1 hour per batch. Testes in the batch
635 were then together transferred to a 1.7 mL Eppendorf tube containing 700 uL 1X PBS. The 1X
636 PBS was then immediately removed and the testes snap-frozen in liquid nitrogen and stored at -
637 80 °C. RNA was extracted from ~150 testes pairs using the RNAeasy Mini kit (Qiagen, 74104).
638 Frozen tissue was dissociated using a 1mL syringe with a 27G needle aspirating up and down
639 ~10 times in RNAeasy Mini kit lysis buffer supplied with 1:100 β-mercaptoethanol as suggested
640 by the kit manual. The quality of extracted RNA was assessed via bioanalyzer, and the
641 concentration measured by Nanodrop and Qubit RNA HS Assay Kit (Life Technologies, Q32852).
642 3' end RNAseq libraries were prepared using the Quant Seq (FWD) kit from Lexogen, using 500
643 ng of total extracted RNA per replicate. The quality of DNA libraries was checked via bioanalyzer,
644 and the concentration measured with Qubit dsDNA HS Assay Kit (Life Technologies, Q32854).

645 Sequencing was performed by the Stanford Genomics facility on an Illumina NextSeq 500
646 platform as in Berry *et al.* (2022). Each biological sample was performed in duplicate. Reads
647 were filtered to retain only reads with 10 consecutive or more 'A'. Reads were then trimmed to
648 remove adaptor sequences and aligned to the *Drosophila* dm6 genome using bowtie 0.12.8
649 with a value of e=5000 to allow for the expected mismatch in the PolyA tail sequence, as in
650 Berry *et al.* (2022). Quantification of PAS usage for the ~500 genes identified in Berry *et al.* 2022
651 was performed using PolyA miner (Yalamanchili *et al.* 2020), using as input a bed file containing
652 the proximal and distal PAS of the ~500 genes identified by Berry *et al.* (2022) as cleaved at a
653 proximal PAS in spermatocytes and a distal PAS in spermatogonia. To consider various
654 spermatocyte stages we generated a master bed file by merging the proximal and distal
655 cleavage sites from 3 different lists: $bam^{-/-}$ (enriched for spermatogonia) vs. 32H (very early
656 spermatocytes), $bam^{-/-}$ vs. 48H (polar spermatocytes) and $bam^{-/-}$ vs 72H (mid-stage maturing
657 spermatocytes). 32H, 48H and 72H indicate hours after heat shock, which causes a pulse of bam
658 expression in $bam^{-/-}$ flies, inducing differentiation, as explained in Kim *et al.* (2017). As in these
659 lists the coordinates of the cleavage sites did not match perfectly, some genes had more than 2
660 cleavage sites considered. However, PolyA miner was devleoped to consider genes with multiple
661 polyA sites.
662
663 ***De novo* PAS calling and APA analysis**.
664 For analysis of changes in PAS usage genome wide between wild-type and knock down samples,
665 we adapted the analysis pipeline from Zeng *et al.* 2024. The sequenced 3' end-seq libraries were
666 adapter-trimmed, and quality filtered using bbduk. The filtered reads were mapped to the dm6
667 *Drosophila* genome using STAR, then extracted for unique-mapping reads using Samtools. If a
668 read was mapped to a region that was immediately upstream of six consecutive genomically
669 encoded As or of a 10-bp region with at least 60% As, it was considered mis-primed and
670 removed. The resulting filtered reads were used for *de novo* PAS calling using LAPA[61] with the
671 default setting, except requiring a replication rate cutoff at 0.95. For each gene, lengthening or
672 shortening was quantified by considering the ratio between the 2 highest changing PAS: the PAS
673 that is more upregulated in the test sample versus control sample and the PAS that is more

674 downregulated in the test sample versus control sample. For this analysis we considered only
675 genes for which both PASs located in the 3'UTR.
676
677 **RNAseq and analysis.**
678 The same total RNA extractions utilized for 3'seq was also used for preparation of short read
679 RNAseq libraries. 1 ug of total extracted RNA was used to prepare libraries following the
680 protocol from the NEBNext Ultra II Direction RNA Library Prep Kit for Illumina, following the
681 section for PolyA selected RNA library preparation. Each biological sample was performed in
682 duplicate. The quality and concentration of the libraries were measured via Bioanalyzer and
683 Qubit, as for 3'seq libraries. Sequencing was performed by Novagene on a NOVAseq, PE 150
684 Illumina Platform. Raw reads were trimmed to remove the Illumina adapters using trim_galore,
685 then aligned to the *Drosophila* dm6 genome using STAR. Differential expression analysis was
686 performed using the package DEseq2.
687
688 **Fly Cell Atlas single nuclear RNAseq analysis**
689 Using the Fly Cell Atlas "Testis" polydT-primed 10x single-nuclear RNA sequencing dataset (Li *et*
690 *al.* 2023), for each of the 111 Leiden 6.0 clusters (as in Berry *et al.* 2022) we aggregated all reads
691 from nuclei in that cluster. For each cluster, reads that uniquely mapped to one of the isoform
692 specific *PCF11* 3'UTRs were counted and normalized by the length of that 3'UTR as well as the
693 total number of reads derived from that cluster. This produced an isoform-specific rpkm value
694 for each Leiden 6.0 cluster that was then log-normalized and plotted using Seurat. Gene
695 expression and cell type annotation plots across all the adult *Drosophila* tissues analyzed in the
696 Fly Cell Atlas project were generated in Seurat using published code modified to visualize
697 relevant genes and cell identities of interest (Raz *et al.* 2023).
698
699 **Western Blots.**
700 Testes from 20-50 (depending on the genotype and the protein needed to be detected) 0-3 days
701 old males were quickly dissected into 1X PBS in a cyclops dissecting dish at room temperature,
702 transferred to a 1.7 mL Eppendorf tube containing 700 uL 1X PBS. The 1X PBS was then removed

703 and the tube snap frozen in liquid nitrogen and stored at -80 °C until needed. Frozen tissue was
704 lysed in 40 uL of cold Lysis Buffer (either RIPA buffer [50 mM TRIS pH 7.5, 250 mM NaCl, 0.1%
705 SDS, 0.5% Sodium deoxycholate to which was added one tablet per 50 uL of protease inhibitor
706 cocktail (cOmplete ULTRA Tablets, Mini, EDTA-free, EASY pack, Roche, REF 05 892 791 001)] or
707 JUME Buffer [10 mM MOPS, 10 mM EDTA, 8M UREA, 1% SDS]). Tissue was dissociated using a 1
708 mL syringe with a 27G needle aspirating up and down ~10 times, then incubated rocking at 4°C
709 for 20 minutes. Following incubation, lysate was centrifuged at 15000 rpm for 5 minutes at 4°C.
710 The supernatant was transferred to a new tube, mixed with 4X Laemli Buffer with β-
711 mercaptoethanol, boiled for 7 minutes at 85 °C, then run on a Mini-PROTEAN TGX Precast
712 Protein Gel (4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 10-well, 50 µl #4561084). The
713 separated proteins were transferred from the gel to a PVDF membrane (BIO-RAD, Cat.#
714 1620177) overnight at 4°C. The following day, the membrane was blocked overnight at 4°C in
715 5% nonfat milk powder dissolved in PBST (1X PBS with 0.1% Tween-20). After blocking, the
716 membrane was incubated for 1 hour at room temperature with primary antibody mix in 2%
717 nonfat milk powder dissolved in PBST, washed 3 times 20 minutes each with PBST at room
718 temperature, incubated with secondary antibody mix for one hour at room temperature in 2%
719 nonfat milk powder dissolved in PBST and washed with PBST again 3 times 20 minutes each at
720 room temperature. Bound antibodies were then revealed using Revvity Health Sciences Inc
721 Western Lightning Plus-ECL, Enhanced Chemiluminescence Substrate (PerkinElmer
722 #NEL104001EA). Primary antibodies used for Western blots: Rabbit anti-HA (1:5000 or 1:1000,
723 Cell Signaling Technologies, #3724), Rabbit anti-PCF11 (1:10.000, kindly gifted from David Scott
724 Gilmour (Zhang and Gilmour 2006)), mouse anti-V5 (1:1000, Invitrogen #46-0705). Secondary
725 antibodies used for Western blots: Peroxidase AffiniPure™ Goat Anti-Rabbit IgG (H+L) (1:10.000,
726 Jackson ImmunoResearch AB_2313567), Peroxidase AffiniPure™ Goat Anti-Mouse IgG (H+L)
727 (1:10.000, Jackson ImmunoResearch AB_10015289).
728
729 **Immunoprecipitation (IP).**
730 For Immunoprecipitation (IP) and co-IP experiments, testes were dissected into 1X PBS in a
731 cyclops dissecting dish at room temperature in batches of 50 flies for a maximum of 30 minutes

732 per batch, transferred into a 1.7 mL Eppendorf tube with 700 uL 1X PBS, the 1X PBS was
733 removed and the tube snap-frozen in liquid nitrogen and stored at -80 °C. For one sample ~100
734 pairs of testes were used. The day before the actual IP, 2 uL per sample of rabbit anti-HA
735 antibodies (Cell Signaling Technologies, #3724) or 2uL of mouse anti-V5 antibodies (Invitrogen
736 #46-0705), were crosslinked to magnetic beads (Dynabeads™ Protein A Invitrogen #10002D for
737 antibodies raised in rabbit, and Dynabeads™ Pan Mouse IgG Invitrogen #11041 for antibodies
738 raised in mouse) as described in Baker *et al.* (2023). Throughout this procedure, each of the
739 following washes of the magnetic beads is done by gently resuspending beads in solution,
740 placing the 1.7 uL containing beads Eppendorf tubes on a magnetic rack and removing the
741 liquid. The day of the experiment, beads crosslinked with antibody were rinsed 1 time in 200
742 mM glycine pH 2.5 and washed once for 5 minutes in 200 mM glycine pH 2.5 to remove non-
743 crosslinked antibodies, then blocked for 1 hour in 10% BSA in 50 mM Tris pH 8, and re-
744 suspended in 20 uL of RIPA buffer and stored at 4 °C while the testis lysate was prepared. Frozen
745 testis tissue was dissociated in RIPA buffer using a 1mL syringe and a 27G needle as in Western
746 Blots experiments. The lysate was rocked at 4°C for 20 minutes, then centrifuged 5 minutes at
747 max speed in a 1.7 mL Eppendorf tube at 4°C to separate lysate from tissue debris. The
748 supernatant was transferred to a new 1.7 mL Eppendorf tube and incubated 45 minutes at 4 °C
749 with beads with no antibody. Tubes containing lysate and beads were placed on a magnetic rack
750 and 40 uL of precleared lysate was saved as input sample. Volume was brought back to 220 uL
751 by adding 40 uL of cold RIPA buffer, pipetted into the tubes containing beads crosslinked to
752 antibody and left incubating for 4 hours at 4°C with rocking. Since we noticed that HA:Cbc in
753 testis lysates tended to form aggregates (see SOM Figure 11), we protein-precipitated (see
754 protocol below) the input sample and resuspended the formed protein-pellet in JUME Buffer. 4X
755 Laemli Sample Buffer (BIO-RAD #1610747) mixed with β-mercaptoethanol was added to the
756 input sample to a final 1X concentration, incubated 7 minutes at 85 °C and frozen. After 4 hours
757 rocking at 4°C, IP samples were placed on a magnet to separate beads from lysate and 40 uL of
758 the supernatant were saved as supernatant sample, mixed with 4X Laemli Sample Buffer with β-
759 mercaptoethanol incubated 7 minutes at 85°C and frozen. The rest of the supernatant was
760 discarded. The beads with attached protein were immediately resuspended in RIPA buffer and

761 washed with RIPA buffer 2 times for 5 minutes rocking at room temperature. Tubes were placed
762 on a magnet to separate beads from RIPA buffer, then RIPA buffer was removed, and the beads
763 were resuspended in 40 uL Elution Buffer (1X cOmplete ULTRA Tablets, 1% SDS, 10 mM EDTA, 50
764 mM Tris pH 8), vortexed, and incubated for 30 minutes at 70°C with frequent (every 5 minutes)
765 vortexing or shaking. Tubes were then centrifuged at max speed for 10 seconds at room
766 temperature, then placed on the magnet to separate beads from the eluate. 40 uL of eluate
767 were moved to another 1.7 mL tube (IP-sample) to which was added 4X Laemli Sample Buffer
768 with β-mercaptoethanol, then the sample was incubated 7 minutes at 85 °C and frozen. The
769 beads were then mixed with 40 uL of elution buffer, mixed with 4X Laemli Sample Buffer with β-
770 mercaptoethanol, incubated 7 minutes at 85 °C and frozen. Samples were then run on a Mini-
771 PROTEAN TGX Precast Protein Gel and the protocol for Western Blots above followed. If the
772 same membrane needed to be re-stained with another antibody, the membrane was stripped
773 by incubation at 55 °C for 30 minutes in 62.5 mM Tris pH 6.8, 2% SDS and 100 mM β-
774 mercaptoethanol, washed 3 times for 10 minutes each at room temperature in PBST and either
775 stored in 1X PBS or blocked again before being re-stained as above. Alternatively, for a milder
776 stripping, the membrane was incubated overnight at 4C with 5% nonfat milk in PBST + 0.1 mM
777 sodium azide. The membrane was blocked again, then re-stained as above.
778
779 **Protein precipitation.**
780 To precipitate proteins, 4 volumes of ice-cold methanol were added to 1 volume of testis lysate,
781 vortexed, and 1 volume of chloroform added. The sample was vortexed again, mixed with 3
782 volumes of H_2O, vortexed and centrifuged in 1.7 mL Eppendorf tubes at 4 °C max speed for 5
783 minutes. After centrifugation, 2 liquid layers formed, with proteins at the interface. The upper
784 liquid layer was removed, and 3 volumes of ice-cold methanol added to the sample. After a
785 quick vortex, the sample was centrifuged at 4 °C max speed for 10 minutes to pellet the
786 proteins. After centrifugation all liquid was removed, and the white protein pellet was air-dried
787 and resuspended in 40 uL of JUME buffer.
788

789 **Figure Legends**

790

791 **Figure 1. Expression of Cleavage Factors during *Drosophila* spermatogenesis**. (A) Schematic
792 representation of the cleavage machinery, adapted from Gallicchio *et al.* (2023). Grey: CPSF =
793 Cleavage and Polyadenylation Factor complex. Orange: CstF = Cleavage and Stimulation Factor
794 complex. Blue: CFI = Cleavage Factor I. Green: CFII = Cleavage Factor II. PAS = Polyadenylation
795 signal. (B) Diagram of a wild-type *Drosophila* testes showing key stages of germ cell
796 differentiation. (C) Diagram of developmentally regulated APA in the *Drosophila* testis resulting
797 in expression of transcripts with long 3'UTRs in spermatogonia but short 3'UTRs in
798 spermatocytes from ~500 genes. (D) Quantification of relative immunofluorescence signal per
799 nuclear area from (E-J). Each dot corresponds to a single nucleus. For each testis, 20
800 spermatogonial nuclei and 20 spermatocyte nuclei were quantified and N=5 testes were
801 quantified per staining, for a total of 100 spermatogonia and 100 spermatocytes. Statistical
802 significance calculated using the Kolmogorov-Smirnov t-test: (****) p-value < 0.0001; (*) p =
803 0.0366. (E-J) Example immunofluorescence images of testis apical regions from samples
804 quantified in D. (Magenta) anti-Vasa marks germ cells. (Cyan) cleavage factor proteins: (E-G, J)
805 Apical region of wild-type (*w1118*) testis stained with anti- (E) CstF50, (F) CstF64, (G) PABP2
806 and (J) PCF11. (H) Apical regions of testes from *CPSF6:GFP* transgenic fly stained with anti-GFP,
807 (I) Apical region of testes from *HA:cbc/HA:cbc* fly stained with anti-HA. Dotted yellow line:
808 border between proliferating spermatogonia (SG) and differentiating spermatocytes (SC). Scale
809 bars = 50 uM.

810

811 **Figure 2. Knock down of *PCF11* and *cbc* in spermatocytes.** (A-C) Apical tips of testes stained
812 with anti-PCF11. (A) Control: *bamGal4*, (B) *PCF11-RNAi-VDRC-103710;bamGal4,* (C) *PCF11-RNAi*
813 *NIG-10228R-4;bamGal4*. (D-E-F) Phase contrast microscope images of whole mount live testes
814 from *Drosophila* lines in (A, B, C) respectively. (G-I) Apical tips of testes stained with anti-HA. (G)
815 Control *HA:cbc/Cyo;bamGal4*, (H) *HA:cbc/+,cbc-RNAi-BL-77426;bamGal4* and (I) *HA:cbc/+,cbc-*
816 *RNAi-VDRC-100686;bamGal4*. (J-L) Phase contrast microscope images of whole mount live
817 testes. (J) Control: *bamGal4*, (K) *cbc-RNAi-BL-77426;bamGal4* and (L) *cbc-RNAi-VDRC-*

818 *100686;bamGal4*. Arrow heads indicate somatic cyst cell nuclei. Scale bars are 50 uM (A-C, G-I)
819 and 200 uM (D-F, J-L). Note that in all the immunofluorescence experiments on *HA:cbc/+* flies,
820 the level of immunofluorescence signal with anti-HA in spermatogonia was below the detection
821 limit for the experimental conditions. In contrast, for the testes quantified in Fig.1-I, which were
822 from homozygous *HA:cbc/HA:cbc* flies, the level of HA:Cbc in nuclei of spermatogonia was high
823 enough to be detected (for comparison between homozygous and heterozygous testes see SOM
824 Figure 1).
825
826 **Figure 3. Knock down of either *PCF11* or *cbc* reduces cleavage at the proximal PAS in**
827 **spermatocytes**. (A) Diagram illustrating the design of the two probe sets. Light blue: probe
828 against the protein coding sequence (CDS), orange: probe against the 3'UTR extension. (B-D'')
829 Confocal microscope images of (B, B'') *bamGal4* (control), (C, C'') *PCF11-RNAi-VDRC-*
830 *103710;bamGal4* and (D, D'') *cbc-RNAi-BL-77426;bamGal4* testes hybridized with probes
831 recognizing the CDS of the *Kpc2* transcripts (cyan) and DAPI (B, C, D), and with probes
832 recognizing the 3'UTR extension of the *Kpc2* transcripts (in orange) (B'', C'', D''). (B'-D', B'''-D''')
833 Zooms in spermatocytes of corresponding regions from (B-D, B''-D'') respectively. (E) IGV plots
834 of 3'seq pile up of reads for the gene *Kpc2* from a sample enriched for SG = spermatogonia (16H
835 PHS = post heat shock, Kim *et al.* 2017; Berry *et al.* 2022) and a sample enriched for
836 spermatocytes (SC = spermatocytes, 72H PHS), to show the APA switch from the distal PAS
837 (dPAS) used in spermatogonia and the proximal PAS (pPAS) used in spermatocytes. (F)
838 Quantification of fluorescence ratio from probes recognizing the *Kpc2* 3'UTR extension (orange)
839 compared to probes recognizing the *Kpc2* CDS (cyan) in spermatocytes. For each genotype N=3
840 testes were quantified. For each image 5 measurements were taken in different spermatocytes
841 areas, which are plotted on the graph. (G-K) Analogous to (A-F) for the *Red gene*. Scale bars are
842 150 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',
843 I'''). p values are (****) = $p<0.0001$ and (***) = $p = 0.0001$. For statistical analysis we used the
844 Welch's t-test.
845

846 **Figure 4. 3' end RNAseq comparing effects on APA of knock downs in spermatocytes of**
847 **different components of the cleavage machinery.** (A, B) Integrative Genomics Viewer (IGV)
848 plots of 3' end RNA sequencing reads for regions encoding the 3'UTR for the (A) *Kpc2* and (B)
849 *Red* genes. Regions of peaks that change in color show mismatches between the PolyA tail and
850 the genome to highlight the 3' end of the mRNA. (C) Heatmap of PolyA index quantified with
851 PolyA Miner across 3'seq libraries. Orange indicates increased cleavage at the distal PAS in knock
852 downs compared to controls, and blue indicates increased cleavage at the proximal PAS in knock
853 downs compared to control. Each row represents one of the >500 genes that switch to a
854 proximal PAS in wild-type spermatocytes compared to spermatogonia from Berry *et al.* (2022).
855 Each column represents a component of the cleavage machinery that was knocked down in
856 spermatocytes. The interval chosen to plot the PolyA index (-3,3) does not represent the
857 maximum and minimum PolyA index observed. In this heatmap genes that showed a statistically
858 non-significant PolyA index in all the samples or a negative PolyA index in all the samples were
859 removed, and in all other cases PolyA indexes that were non-significant were put equal to zero
860 (another heatmap with these genes showed is reported in SOM Figure 2) (D, E) Integrative
861 Genomics Viewer (IGV) plots of 3' end RNA sequencing libraries for the genes (D) *oaf* and (E)
862 *nudE*. (F-H) Venn diagrams representing number of genes with polyA index > 0.5 in *cbc* knock
863 down libraries (dark green), *PCF11* VDRC knock down libraries (light green), *PABP2* knock down
864 libraries (red) and components of CstF complex (orange).
865
866 **Figure 5. *PCF11* and *cbc* transcripts are both upregulated in spermatocytes with *cbc***
867 **particularly highly expressed in testis compared to other tissues.** (A) Diagram adapted from
868 FlyBase of *PCF11* and *cbc* loci. Light blue boxes represent 5'UTRs, green and grey boxes
869 represent exons and pink boxes represent 3'UTRs. Introns represented with a straight black line.
870 In yellow are represented the transcripts regions targeted by HCR-FISH probes for RNA *in situs*.
871 Green and dotted-dark green boxes above the *PCF11* locus represent respectively the CID
872 domain and a region in which the Cbc binding domain could be located. (B-G) Scope diagrams of
873 Fly Cell Atlas single nuclear RNAseq data (Li *et al.* 2023) showing: (B) *PCF11* and (C) *cbc*
874 expression across all the analyzed adult tissues, (D) Purple marking nuclei from the testis plus

875 seminal vesicle sample across all analyzed adult tissues, (E-G) Umap plots showing (E) *PCF11*
876 and (F) *cbc* expression in the testis plus semnal vesicle sample, (G) orange marking nuclei from
877 the male germ-line clusters. (H-J) Confocal microscope images of apical tips of wild-type
878 (*w1118*) testes hybridized with (H) probes against *PCF11* long isoforms (white) and DAPI (blue),
879 (I) probes against *PCF11*-RB isoform (white) and DAPI (blue), (J) probes against *cbc* (white) and
880 DAPI (blue). In (H, I, J) the boxes highlight a region of spermatogonia (SG) zoomed in (H', I', J')
881 and a region of spermatocytes (SC) zoomed in (H'', I'', J''). (K, L) Umap from Fly Cell Atlas single
882 nuclear data from testis plus seminal vesicle sample showing *PCF11-long* (K) and *PCF11-RB* (L)
883 expression across the different clusters. Scalebars are 50 uM in (H, I, J) and 20 uM in (H', H'', I',
884 I'', J', J'').
885
886 **Figure 6. knock down of *PCF11* caused decrease of Cbc protein**. (A) Western Blots of wild-type
887 (*w1118*) testis lysate probes with anti-PCF11. (B) Western Blots of *HA:cbc* homozygous testes
888 lysate probed with anti-HA. (C) Immunoprecipitation (IP) of *HA:cbc* in wild-type (*w1118*) and
889 *HA:cbc* homozygous testes followed by PAGE and western blots probed with anti-PCF11 or anti-
890 HA. Input = 20% of lysate before IP. (D, E) IGV plots of (D) *cbc* and (E) *PCF11* loci of RNAseq data
891 from *bamGal4* (control) samples (in grey), *PCF11-RNAi-VDRC-103710;bamGal4* samples
892 (magenta) and *cbc-RNAi-BL-77426;bamGal4* samples (teal). (F-J) Immunofluorescence images
893 showing apical tips of testes stained with anti-Vasa to mark germ cells, anti-HA, and anti-PCF11.
894 (F-F'') *HA:cbc/CyO;bamGal4* (control), (G-G'') *HA:cbc/+,PCF11-RNAi-VDRC-103710;bamGal4*, (H-
895 H'') *HA:cbc/+,PCF11-RNAi-NIG-10228R-4;bamGal4*, (I-I'') *HA:cbc/+,cbc-RNAi-BL-77426;bamGal4*
896 and (J-J'') *HA:cbc/+,cbc-RNAi-VDRC-100686;bamGal4*. (K, L) Quantification of Fluorescence/Area
897 from (K) HA antibodies (teal) and (L) PCF11 antibodies (magenta). For each testis 10
898 spermatogonial and 10 spermatocyte nuclei were quantified for both anti-PCF11 and anti-HA
899 fluorescence and then averaged. For each genotype N=3 testes were quantified. Statistical
900 analysis performed using the Welch's t-test. p-values: *bamGal4_PCF11* = 0.0011 (**),
901 b*amGal4_cbc* = 0.0005 (***), *PCF11-RNAi-NIG10228R-4_PCF11* = 0.0341 (*), *cbc-RNAi-*
902 *BL77427_PCF11* < 0.0001 (****), *cbc-RNAi-VDRC100686_PCF11* = 0.0303 (*). Scalebars are 50
903 uM (F-J''). Entire gels for western blots and co-IP are reported in SOM Figure 11.

904

905 **Figure 7. Overexpression of PCF11 and Cbc in spermatogonia shifted cleavage towards the**
906 **more proximal site for some transcripts**. (A-D) *bam*$^{-/-}$ testes stained with anti-Vasa, anti-HA and
907 anti-V5. (A-A'') *bam* mutant control lacking overexpression constructs (*nosGal4/Y; +/CyO;*
908 *bam*^{86}*,bamGal4/bam*^{1}*)*, (B-B'') *bam* mutant testes overexpressing *PCF11-RF-HA (nosGal4/Y;*
909 *UAS>PCF11-RF-HA/+; bam*^{86}*,bamGal4/bam*^{1}*),* (C-C'') *bam* mutant testes overexpressing *cbc-V5*
910 *(nosGal4/Y; UAS>cbc-V5/+; bam*^{86}*,bamGal4/bam*^{1}*),* (D-D'') *bam* mutant testes overexpressing
911 both *PCF11-RF-HA* and *cbc-V5 (nosGal4/Y; UAS>PCF11-RF-HA, UAS>cbc-V5/+;*
912 *bam*^{86}*,bamGal4/bam*^{1}*).* (E) Pink: quantification of nuclear/cytoplasmic ratio of PCF11-RF-HA in
913 *bam*$^{-/-}$ testes overexpressing *PCF11-RF-HA* alone or *PCF11-RF-HA* plus *cbc-V5*. Blue:
914 quantification of nuclear/cytoplasmic ratio of cbc-V5 in *bam*$^{-/-}$ testes overexpressing *cbc-V5*
915 alone or *PCF11-RF-HA* plus *cbc-V5*. For each sample N=5 testes were quantified. Statistical
916 analysis was performed using the Welch's t-test. p-values are 0.8103 (ns) and <0.0001 (****).
917 (F) Immunoprecipitation (IP) of *cbc-V5* from lysates of *bam*$^{-/-}$ testes in which either *PCF11-RF-HA*
918 plus *cbc-V5* or *PCF11-RF-HA* alone (control) were overexpressed, followed by PAGE and western
919 blot probed with anti-HA and anti-V5 antibodies. Input = 20% of lysate before IP. (G) Venn
920 diagram showing the number of genes that are affected in opposite APA directions by CFII knock
921 downs (in green) or CFII overexpression (in red). The green circle represents all the genes that
922 had a polyA index > 0.5 (lengthening) in either *PCF11* or *cbc* knock downs, while the red circle
923 all the genes that had a polyA index < -0.5 (shortening) in either *PCF11-RF-HA* overexpression,
924 or *cbc-V5* overexpression, or *PCF11-RF-HA* plus *cbc-V5* overexpression. (H-K) IGV plots showing
925 3' end seq reads from libraries for overexpression and knock down samples for the (H) *Rop,* (I)
926 *orb,* (J) *Mitofilin* and (K) *bor* loci. OE=overexpression, SG=spermatogonia. (CTRL) = *nosGal4/Y;*
927 *+/+; bam*^{86}*,bamGal4/bam*^{1}*,* (*cbc* OE) = *nosGal4/Y; UAS>cbc-V5/+; bam*^{86}*,bamGal4/bam*^{1}*, (PCF11*
928 OE) = *NosGal4/Y; UAS>PCF11-RF-HA/+; bam*^{86}*,bamGal4/bam*^{1}*, (cbc + PCF11* OE) = *nosGal4/Y;*
929 *UAS>PCF11-RF-HA, UAS>cbc-V5/+; bam*^{86}*,bamGal4/bam*^{1}*.*

930 **SOM Figure Legend**

931

932 **SOM 1. Comparison of anti-HA staining between homozygous *HA:cbc/HA:cbc* and**
933 **heterozygous *HA:cbc/CyO* testes.** (A) Apical testis from a male homozygous for *HA:cbc* and (B) a
934 male *HA:cbc/CyO* heterozygous stained with anti-HA (Green) and DAPI (Blue). (A'-B') show the
935 HA:Cbc channel alone in grey corresponding to A and B. Scalebars are 100 uM. Green line marks
936 spermatogonia region.

937

938 **SOM 2.** Heatmap related to Figure 4 showing more of the >500 genes from Berry et al. (2022)
939 compared to the heatmap reported in Figure 4 C. Now are reported also genes that either have
940 a PolyA index equal to zero in all the knock downs, a statistically non-significant PolyA index in
941 all the knock downs, or a negative PolyA index in all the knock downs (that were omitted from
942 the heatmap in Figure 4 C).

943

944 **SOM 3.** Volcano plots showing gene expression levels from RNAseq data from (A) *cbc-RNAi-BL*
945 *77426;bamGal4* and (B) *PCF11-RNAi-VDRC-103710;bamGal4* knock down testes versus
946 *bamGal4* (control). In black are represented the genes that switch PAS between spermatogonia
947 and spermatocytes from Berry *et al.* 2022 and that show a PolyA index either smaller than -0.5
948 or larger than 0.5 in the 3'seq libraries from the corresponding knock down.

949

950 **SOM 4. Verification of *PABP2* knock down efficiency**. (A, B) Apical tips of testes stained with
951 anti-PABP2 antibodies and DAPI. (A) *bamGal4* (control) testes and (B) *PABP2-RNAi-VDRC-*
952 *106466;bamGal4* testes. Arrow heads indicate somatic cyst cell nuclei, and a green line
953 highlights the testis region with spermatogonia. (C-E) Phase contrast images of whole testes
954 from: (C) Control: *bamGal4* testes, (D) *PABP2-RNAi-VDRC-106466;bamGal4* testes and (E)
955 *PABP2-RNAi-VDRC-33499;bamGal4*. Scalebars are 50 uM (A, B) and 200 uM (C, D).

956

957 **SOM 5. Verification of knock down efficiency of CstF components**. Apical tips of testes from:
958 (A) *bamGal4* (control) and (B) *CstF50-RNAi-BL-77377;bamGal4* males stained with anti-CstF50.

959 (C) Phase contrast images of whole testes from *CstF50-RNAi-BL-77377;bamGal4* male. Apical tip
960 of testes from (D) *bamGal4* (control) and (E) *CstF64-RNAi-BL-65987;bamGal4* testes stained
961 with anti-CstF64 antibodies. (F) Phase contrast images of whole *CstF64-RNAi-BL-65987;bamGal4*
962 testis. Arrow heads indicate somatic cyst cell nuclei, and a green line highlights the
963 spermatogonia region. (G) Phase contrast microscopy of *su(f)-RNAi-BL-65693;bamGal4* testis.
964 (H) Phase contrast images of whole *bamGal4* testis. Scalebars are 50 uM (A, B, D, E) and 200 uM
965 (C, F, G, H).
966
967 **SOM 6. Knock down of *su(f)* or *CstF64* in spermatocytes alters CstF50.** (A, B) *bamGal4* (control)
968 testis stained for (A) CstF64, and (B) CstF50. (C, D) *su(f-RNAi-BL-65693;bamGal4* testis and (E, F)
969 *CstF64-RNAi-BL-65987;bamGal4* testis stained for (C, E) CstF64, and (D, F) CstF50. Scalebars are
970 100 uM.
971
972 **SOM 7. Verification of *CPSF6* knock down efficiency**. Phase contrast images of whole testes
973 from (A-A''') *CPSF6-RNAi-BL-34804;bamGal4* males, (B) a *CPSF6-RNAi-VDRC-107147;bamGal4*
974 male and (C) a *bamGal4* male. Scalebars are 200 uM.
975
976 **SOM 8-9**. Changes in 3' end cut site usage in knock down testes vs Control (*bamGal4*) testes on
977 *de novo* mapped PAS sites on annotated 3'UTRs.
978
979 **SOM 10.** Umap from Fly Cell Atlas single nuclear RNA-seq (Raz *et al.* 2023; Li *et al.* 2023)
980 showing relative expression of *PCF11-RC* in the testis and seminal vesicle sample.
981
982 **SOM 11.** Entire gels for Western Blots and co-IP from Figure 6.
983
984 **SOM 12. Cbc:HA is downregulated (but still present) upon *PCF11* knock down.** Apical tips of
985 testes from (A, A') *HA:cbc/HA:cbc;bamGal4* (B, B') *HA:cbc,PCF11-RNAi-VDRC-*
986 *103710/HA:cbc;bamGal4* (C, C') *HA:cbc,PCF11-RNAi-NIG-10228R-4/HA:cbc;bamGal4* stained
987 with antibodies against HA and PCF11. (D, E) Quantification of fluorescent signal from (D) PCF11
988 and (E) HA antibodies respectively in spermatogonia and spermatocytes as quantified in Figure

989 6. P-values are: (D) (*) = 0.0398 and (E) (*) = 0.0343 in *bamGal4* and (*) = 0.0438 in *VDRC*
990 *103710.* Scalebars are 100 uM.
991
992 **SOM 13.** Entire gels from co-IP in Figure 7
993
994 **SOM 14.** Heatmap reporting the PolyA index for genes from Berry et al. (2022) in 3'seq libraries
995 from *bam*$^{-/-}$ testes (filled with spermatogonia) overexpressing either *cbc-V5* alone, or *PCF11-RF-*
996 *HA* alone, or botch *cbc-V5* and *PCF11-RF-HA*. Blue: proximal PAS more regulated in a sample
997 overexpressing either *cbc-V5*, or *PCF11-RF-HA*, or both compared to control *bam*$^{-/-}$ testes means
998 that the percentage of transcripts cleaved at the proximal site typical of spermatocytes was
999 higher ion the overexpression sample compared to *bam*$^{-/-}$ testes in which neither *PCF11-RF-HA*
1000 nor *cbc-V5* was overexpressed.
1001
1002 **SOM 15.** IGV plots of 3'seq reads of *PCF11* locus. Gray arrow marks a PAS very 5' in the locus
1003 that is utilized in samples enriched for spermatocytes but not much in samples enriched for
1004 spermatogonia. Cleavage at this upstream PAS is strongly downregulated in testes where either
1005 *PCF11* or *cbc* were knocked down in spermatocytes. SG = spermatogonia, SC = spermatocytes.

1006 **Acknowledgements**

1007 We thank all the members of the Fuller lab for suggestions and comments on this
1008 manuscript and Dr. Nicole Martinez and Dr. Lauren Goins for their valuable feedback on the
1009 manuscript and suggested experiments. We thank Dr. Catherine Baker for suggestions on
1010 Immunoprecipitations and for proofreading the manuscript, and Benjamin Bolival for helping in
1011 maintaining fly crosses. We thank Dr. Jingxun Chen from the Brunet lab at Stanford for her help
1012 in designing HCR-FISH probes and for sharing some of the reagents, the Wang lab for the *HA:cbc*
1013 CRISPR tag flies, the Lis lab for anti-CstF50, the Simonelig lab for anti-PABP2, the Dominski lab
1014 for anti-CstF64 and the Gilmour lab for anti-PCF11 antibodies. We thank Daniel Mokhtari for his
1015 help parsing the Fly Cell Atlas dataset, and Dr. Hari Krishna Yalamanchili for his help in running
1016 PolyA miner. We thank the Cell Science Imaging facility at Stanford, supported, in part, by Award
1017 Number 1S10OD010580-01A1 from the National Center for Research Resources (NCRR) (The
1018 contents of this manuscript are solely the responsibility of the authors and do not necessarily
1019 represent the official views of the NCRR or the National Institutes of Health), particularly the
1020 CSIF facility manager Kitty Lee. We thank the Stanford Fly Media Center as well and all the staff
1021 working in the Developmental Biology department. L.G. was supported by an AICF (American
1022 Italian Cancer Foundation) postdoctoral fellowship (2021-2023). This research was supported
1023 by National Institutes of Health grant R35 GM136433) and funds from the Katherine D.
1024 McCormick and Stanley McCormick Memorial Professorship and the Reed-Hodgson
1025 Professorship in Human Biology to M.T.F. This manuscript is dedicated to Cuchuflí.

1026

1027 **Author Contributions**

1028 LG and MTF conceived the project and wrote the manuscript. LG designed the
1029 experimental work and performed or supervised all experimental and computational
1030 procedures. NRM contributed to experimental design and bioinformatic analysis. FMP
1031 contributed to experimental design and performed molecular cloning. IN performed
1032 immunofluorescence experiments. SS performed bioinformatic analysis of the Fly Cell Atlas
1033 data. YZ performed bioinformatic analysis to *de novo* map PASs. Funding was acquired by LG
1034 and MTF.

1035 **References**

1036 An, J. J., K. Gharami, G. Y. Liao, N. H. Woo, A. G. Lau *et al.*, 2008 Distinct Role of Long 3′ UTR
1037 BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. Cell.

1038 Baker, C. C., L. Gallicchio, N. R. Matias, D. F. Porter, L. Parsanian *et al.*, 2023 Cell-type-specific
1039 interacting proteins collaborate to regulate the timing of Cyclin B protein expression in
1040 male meiotic prophase. Development dev.201709.

1041 Bedbrook, C. N., R. D. Nath, R. Nagvekar, K. Deisseroth, and A. Brunet, 2023 Rapid and precise
1042 genome engineering in a naturally short-lived vertebrate (L. Moons, D. Y. R. Stainier, L.
1043 Moons, & P. Abitua, Eds.). Elife 12: e80639.

1044 Benoit, B., M. Simonelig, A. Nemeth, N. Aulner, U. Kühn *et al.*, 1999 The Drosophila poly(A)-
1045 binding protein II is ubiquitous throughout Drosophila development and has the same
1046 function in mRNA polyadenylation as its bovine homolog in vitro. Nucleic Acids Res. 27:
1047 3771–3778.

1048 Berry, C. W., G. H. Olivares, L. Gallicchio, G. Ramaswami, A. Glavic *et al.*, 2022 Developmentally
1049 regulated alternate 3′ end cleavage of nascent transcripts controls dynamic changes in
1050 protein expression in an adult stem cell lineage. Genes Dev. 36: 916–935.

1051 Chen, D., and D. M. McKearin, 2003 A discrete transcriptional silencer in the bam gene
1052 determines asymmetric division of the Drosophila germline stem cell. Development 130:
1053 1159–1170.

1054 Choi, H. M. T., M. Schwarzkopf, M. E. Fornace, A. Acharya, G. Artavanis *et al.*, 2018 Third-
1055 generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive,
1056 versatile, robust. Development 145: dev165753.

1057 Clerici, M., M. Faini, L. M. Muckenfuss, R. Aebersold, and M. Jinek, 2018 Structural basis of
1058 AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat. Struct. Mol.
1059 Biol. 25: 135–138.

1060 Doren, M. Van, A. L. Williamson, and R. Lehmann, 1998 Regulation of zygotic gene expression in
1061 Drosophila primordial germ cells. Curr. Biol. 8: 243–246.

1062 Gallicchio, L., G. H. Olivares, C. W. Berry, and M. T. Fuller, 2023 Regulation and function of
1063 alternative polyadenylation in development and differentiation. RNA Biol. 20: 908–925.

1064 Grassi, E., R. Santoro, A. Umbach, A. Grosso, S. Oliviero *et al.*, 2019 Choice of alternative
1065 polyadenylation sites, mediated by the rna-binding protein Elavl3, plays a role in
1066 differentiation of inhibitory neuronal progenitors. Front. Cell. Neurosci.
1067 Grozdanov, P. N., E. Masoumzadeh, M. P. Latham, and C. C. MacDonald, 2018 The structural
1068 basis of CstF-77 modulation of cleavage and polyadenylation through stimulation of CstF-
1069 64 activity. Nucleic Acids Res. 46: 12022–12039.
1070 Gruber, A. J., and M. Zavolan, 2019 Alternative cleavage and polyadenylation in health and
1071 disease. Nat. Rev. Genet.
1072 Guéguéniat, J., A. F. Dupin, J. Stojko, L. Beaurepaire, S. Cianférani *et al.*, 2017 Distinct roles of
1073 PCF11 zinc-binding domains in pre-mRNA 3-end processing. Nucleic Acids Res.
1074 Hilgers, V., S. B. Lemke, and M. Levine, 2012 ELAV mediates 3′ UTR extension in the Drosophila
1075 nervous system. Genes Dev. 26: 2259–2264.
1076 Jing, Z., H. Linda, and M. Claire, 1999 Formation of mRNA 3′ Ends in Eukaryotes: Mechanism,
1077 Regulation, and Interrelationships with Other Steps in mRNA Synthesis. Microbiol. Mol.
1078 Biol. Rev. 63: 405–445.
1079 Jongmin Kim, Chenggang Lu, Shrividhya Srinivasan, Stephan Awe, Alexander Brehm *et al.*, 2017
1080 Blocking promiscuous activation at cryptic promoters directs cell type – specific gene
1081 expression. Science (80-.).
1082 Kamieniarz-Gdula, K., M. R. Gdula, K. Panser, T. Nojima, J. Monks *et al.*, 2019 Selective Roles of
1083 Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Mol. Cell 74: 158-
1084 172.e9.
1085 Lee, S., Y.-C. Chen, A. E. Gillen, J. M. Taliaferro, B. Deplancke *et al.*, 2022 Diverse cell-specific
1086 patterns of alternative polyadenylation in Drosophila. Nat. Commun. 13: 5372.
1087 Li, H., J. Janssens, M. De Waegeneer, S. S. Kolluru, K. Davie *et al.*, 2023 Fly Cell Atlas: A single-
1088 nucleus transcriptomic atlas of the adult fruit fly. Science (80-.). 375: eabk2432.
1089 Li, W., J. Y. Park, D. Zheng, M. Hoque, G. Yehia *et al.*, 2016 Alternative cleavage and
1090 polyadenylation in spermatogenesis connects chromatin regulation with post-
1091 transcriptional control. BMC Biol.
1092 Li, W., B. You, M. Hoque, D. Zheng, W. Luo *et al.*, 2015 Systematic Profiling of Poly(A)+

1093 Transcripts Modulated by Core 3’ End Processing and Splicing Factors Reveals Regulatory
1094 Rules of Alternative Cleavage and Polyadenylation. PLOS Genet. 11: e1005166.
1095 MacDonald, C. C., J. Wilusz, and T. Shenk, 1994 The 64-Kilodalton Subunit of the CstF
1096 Polyadenylation Factor Binds to Pre-mRNAs Downstream of the Cleavage Site and
1097 Influences Cleavage Site Location. Mol. Cell. Biol. 14: 6647–6654.
1098 Mansfield, K. D., and J. D. Keene, 2012 Neuron-specific ELAV/Hu proteins suppress HuR mRNA
1099 during neuronal differentiation by alternative polyadenylation. Nucleic Acids Res. 40:
1100 2734–2746.
1101 Marzluff, W. F., E. J. Wagner, and R. J. Duronio, 2008 Metabolism and regulation of canonical
1102 histone mRNAs: Life without a poly(A) tail. Nat. Rev. Genet.
1103 Mayr, C., and D. P. Bartel, 2009 Widespread Shortening of 3′UTRs by Alternative Cleavage and
1104 Polyadenylation Activates Oncogenes in Cancer Cells. Cell 138: 673–684.
1105 McKearin, D., and B. Ohlstein, 1995 A role for the Drosophila Bag-of-marbles protein in the
1106 differentiation of cystoblasts from germline stem cells. Development 121: 2937–2947.
1107 McKearin, D. M., and A. C. Spradling, 1990 bag-of-marbles: a Drosophila gene required to
1108 initiate both male and female gametogenesis. Genes Dev. 4: 2242–2251.
1109 Miura, P., S. Shenker, C. Andreu-Agullo, J. O. Westholm, and E. C. Lai, 2013 Widespread and
1110 extensive lengthening of 39 UTRs in the mammalian brain. Genome Res.
1111 Morris, A. R., A. Bos, B. Diosdado, K. Rooijers, R. Elkon *et al.*, 2012 Alternative cleavage and
1112 polyadenylation during colorectal cancer development. Clin. Cancer Res. 18: 5256–5266.
1113 Ni, Z., A. Saunders, N. J. Fuda, J. Yao, J.-R. Suarez *et al.*, 2008 P-TEFb Is Critical for the
1114 Maturation of RNA Polymerase II into Productive Elongation In Vivo. Mol. Cell. Biol. 28:
1115 1161–1170.
1116 Ogorodnikov, A., M. Levin, S. Tattikota, S. Tokalov, M. Hoque *et al.*, 2018 Transcriptome 3′end
1117 organization by PCF11 links alternative polyadenylation to formation and neuronal
1118 differentiation of neuroblastoma. Nat. Commun.
1119 Raz, A. A., G. S. Vida, S. R. Stern, S. Mahadevaraju, J. M. Fingerhut *et al.*, 2023 Emergent
1120 dynamics of adult stem cell lineages from single nucleus and single cell RNA-Seq of
1121 Drosophila testes (M. Buszczak, U. Banerjee, M. Buszczak, & T. G. Nystul, Eds.). Elife 12:

1122 e82201.
1123 Sadowski, M., B. Dichtl, W. Hübner, and W. Keller, 2003 Independent functions of yeast Pcf11p
1124 in pre-mRNA 3′ end processing and in transcription termination. EMBO J. 22: 2167–2177.
1125 Sarov, M., C. Barz, H. Jambor, M. Y. Hein, C. Schmied *et al.*, 2016 A genome-wide resource for
1126 the analysis of protein localisation in Drosophila. Elife 5:.
1127 Schäfer, P., C. Tüting, L. Schönemann, U. Kühn, T. Treiber *et al.*, 2018 Reconstitution of
1128 mammalian cleavage factor II involved in 3′ processing of mRNA precursors. RNA.
1129 Schulte, S. J., M. E. Fornace, J. K. Hall, G. J. Shin, and N. A. Pierce, 2024 HCR spectral imaging:
1130 10-plex, quantitative, high-resolution RNA and protein imaging in highly autofluorescent
1131 samples. Development 151: dev202307.
1132 Shan, L., C. Wu, D. Chen, L. Hou, X. Li *et al.*, 2017 Regulators of alternative polyadenylation
1133 operate at the transition from mitosis to meiosis. J. Genet. Genomics 44: 95–106.
1134 Skrajna, A., X. Yang, M. Dadlez, W. F. Marzluff, and Z. Dominski, 2018 Protein composition of
1135 catalytically active U7-dependent processing complexes assembled on histone pre-mRNA
1136 containing biotin and a photo-cleavable linker. Nucleic Acids Res. 46: 4752–4770.
1137 Tian, B., and J. L. Manley, 2016 Alternative polyadenylation of mRNA precursors. Nat. Rev. Mol.
1138 Cell Biol.
1139 Trivedi, V., H. M. T. Choi, S. E. Fraser, and N. A. Pierce, 2018 Multidimensional quantitative
1140 analysis of mRNA expression within intact vertebrate embryos. Development 145:
1141 dev156869.
1142 Turner, R. E., L. M. Henneken, M. Liem-Weits, P. F. Harrison, A. Swaminathan *et al.*, 2020
1143 Requirement for cleavage factor IIm in the control of alternative polyadenylation in breast
1144 cancer cells. RNA 26: 969–981.
1145 Vallejos Baier, R., J. Picao-Osorio, and C. R. Alonso, 2017 Molecular Regulation of Alternative
1146 Polyadenylation (APA) within the Drosophila Nervous System. J. Mol. Biol.
1147 Venkat, S., A. A. Tisdale, J. R. Schwarz, A. A. Alahmari, H. Carlo Maurer *et al.*, 2020 Alternative
1148 polyadenylation drives oncogenic gene expression in pancreatic ductal adenocarcinoma.
1149 Genome Res. 30: 347–360.
1150 Wu, J., X. Li, Z. Gao, L. Pang, X. Liu *et al.*, 2021 RNA kinase CLP1/Cbc regulates meiosis initiation

1151 in spermatogenesis. Hum. Mol. Genet. 30: 1569–1578.
1152 Yalamanchili, H. K., C. E. Alcott, P. Ji, E. J. Wagner, H. Y. Zoghbi *et al.*, 2020 PolyA-miner:
1153 Accurate assessment of differential alternative poly-adenylation from 3Seq data using
1154 vector projections and non-negative matrix factorization. Nucleic Acids Res.
1155 Yang, Q., G. M. Gilmartin, and S. Doublié, 2011 The structure of human Cleavage Factor Im hints
1156 at functions beyond UGUA-specific RNA binding. RNA Biol. 8: 748–753.
1157 Zeng, Y., A. Lovchykova, T. Akiyama, C. Liu, C. Guo *et al.*, 2024 TDP-43 nuclear loss in FTD/ALS
1158 causes widespread alternative polyadenylation changes. bioRxiv 2024.01.22.575730.
1159 Zhang, Z., and D. S. Gilmour, 2006 Pcf11 Is a Termination Factor in Drosophila that Dismantles
1160 the Elongation Complex by Bridging the CTD of RNA Polymerase II to the Nascent
1161 Transcript. Mol. Cell 21: 65–74.
1162 Zhu, Y., X. Wang, E. Forouzmand, J. Jeong, F. Qiao *et al.*, 2018 Molecular Mechanisms for CFIm-
1163 Mediated Regulation of mRNA Alternative Polyadenylation. Mol. Cell 69: 62-74.e4.
1164
1165
1166

Fig.1

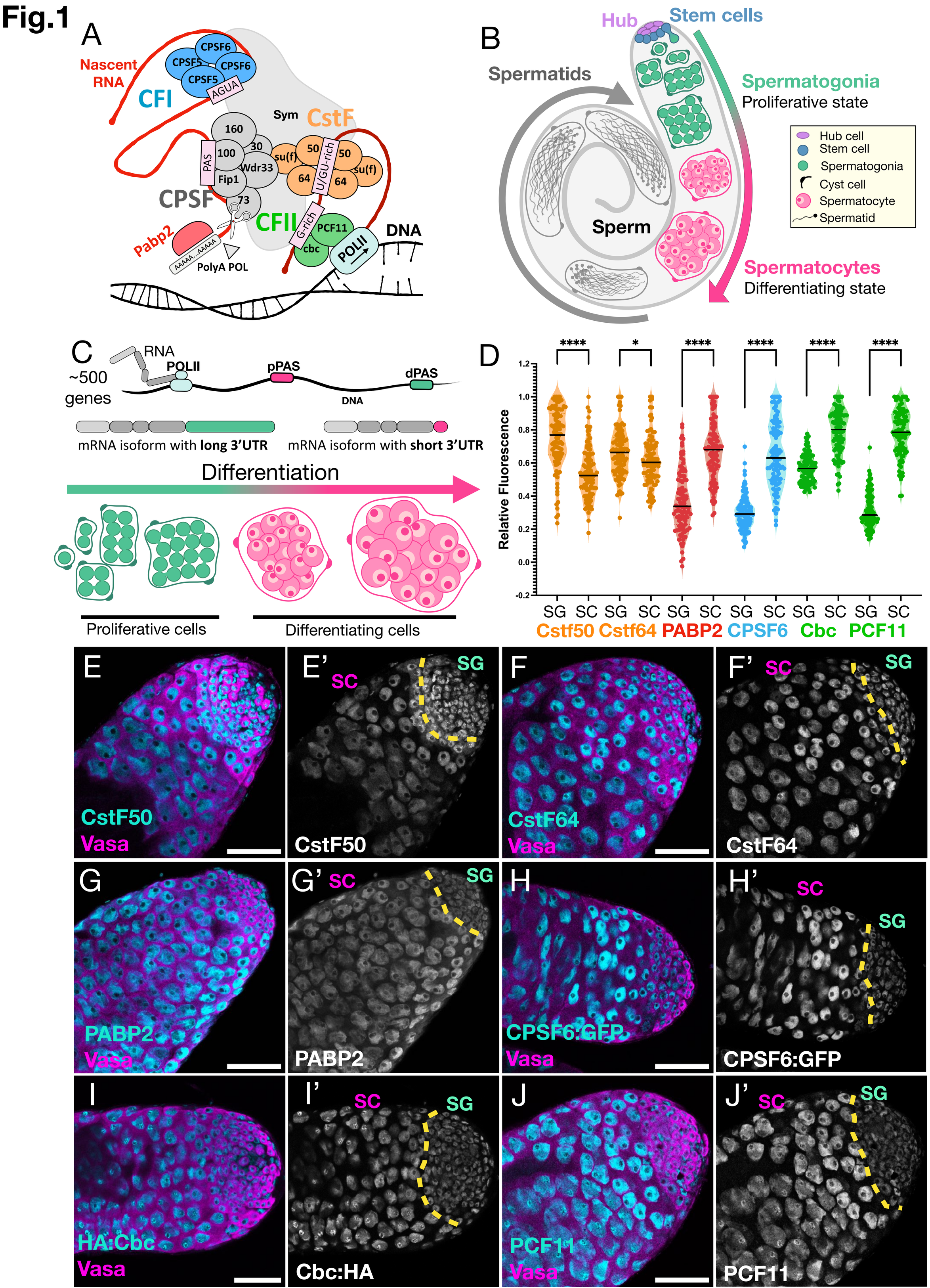

A
Nascent RNA
CPSF6
CPSF5
CPSF6
CPSF5
AGUA
CFI
Sym
CstF
160
100
30
Wdr33
Fip1
PAS
su(f)
50
64
U/GU-rich
50
64
su(f)
CPSF
73
CFII
G-rich
PCF11
cbc
POLII
DNA
Pabp2
AAAAA...AAAAA
PolyA POL
B
Hub
Stem cells
Spermatids
Spermatogonia
Proliferative state
Hub cell
Stem cell
Spermatogonia
Cyst cell
Spermatocyte
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
Cstf50 Cstf64 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

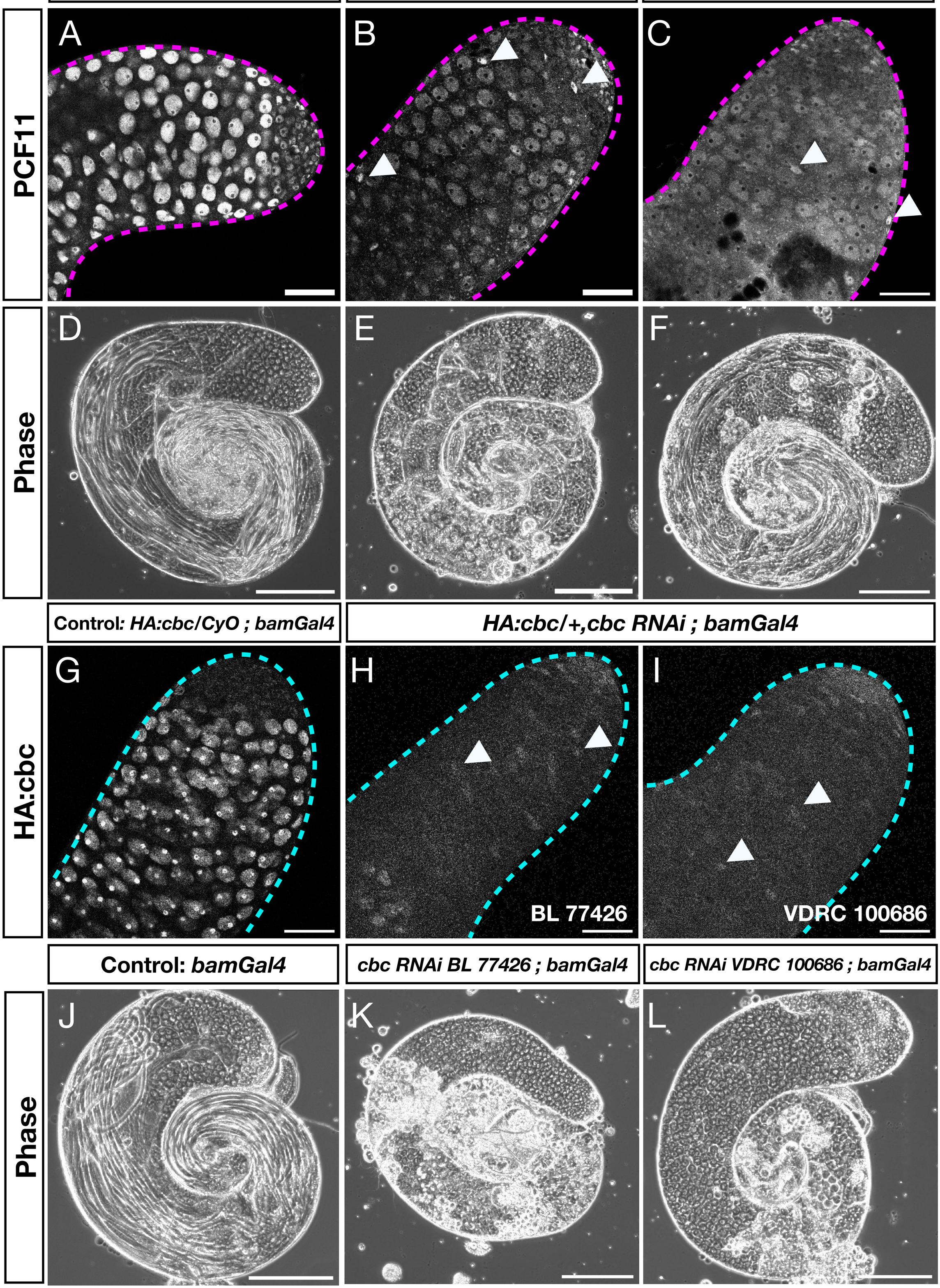

Control: bamGal4
PCF11 RNAi VDRC 103710 ; bamGal4
PCF11 RNAi NIG 10228R-4 ; bamGal4
PCF11
A
B
C
Phase
D
E
F
Control: HA:cbc/CyO ; bamGal4
HA:cbc/+,cbc RNAi ; bamGal4
HA:cbc
G
H
I
BL 77426
VDRC 100686
Control: bamGal4
cbc RNAi BL 77426 ; bamGal4
cbc RNAi VDRC 100686 ; bamGal4
Phase
J
K
L

Fig.3

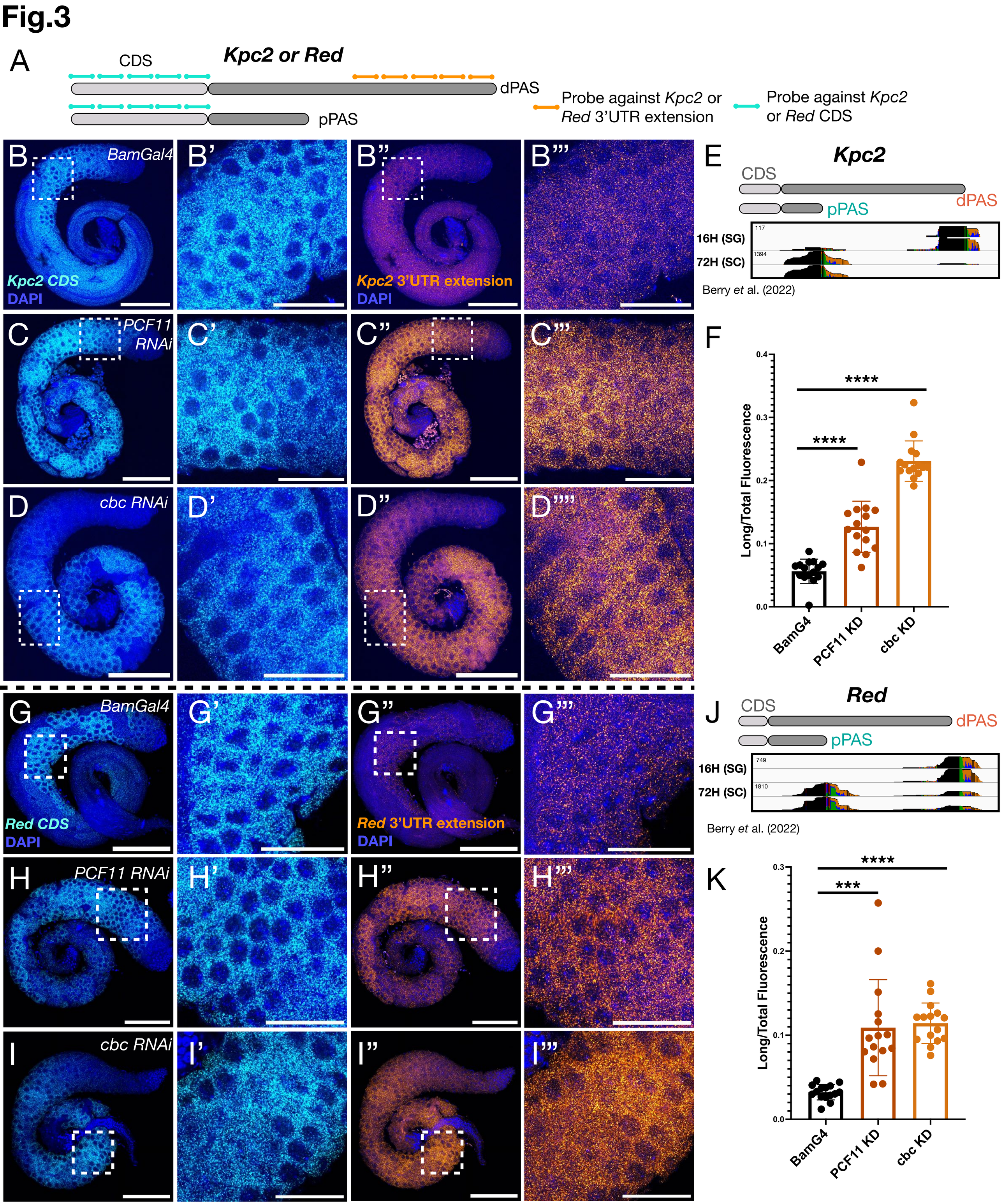

A
CDS
Kpc2 or Red
dPAS
pPAS
Probe against Kpc2 or Red 3'UTR extension
Probe against Kpc2 or Red CDS
B
BamGal4
Kpc2 CDS
DAPI
B'
B''
Kpc2 3'UTR extension
DAPI
B'''
C
PCF11 RNAi
C'
C''
C'''
D
cbc RNAi
D'
D''
D''''
E
Kpc2
CDS
dPAS
pPAS
16H (SG)
72H (SC)
117
1394
Berry et al. (2022)
F
Long/Total Fluorescence
0.4
0.3
0.2
0.1
0.0
****
****
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)
749
1810
Berry et al. (2022)
K
Long/Total Fluorescence
0.3
0.2
0.1
0.0
****
***
BamG4
PCF11 KD
cbc KD

Fig.4

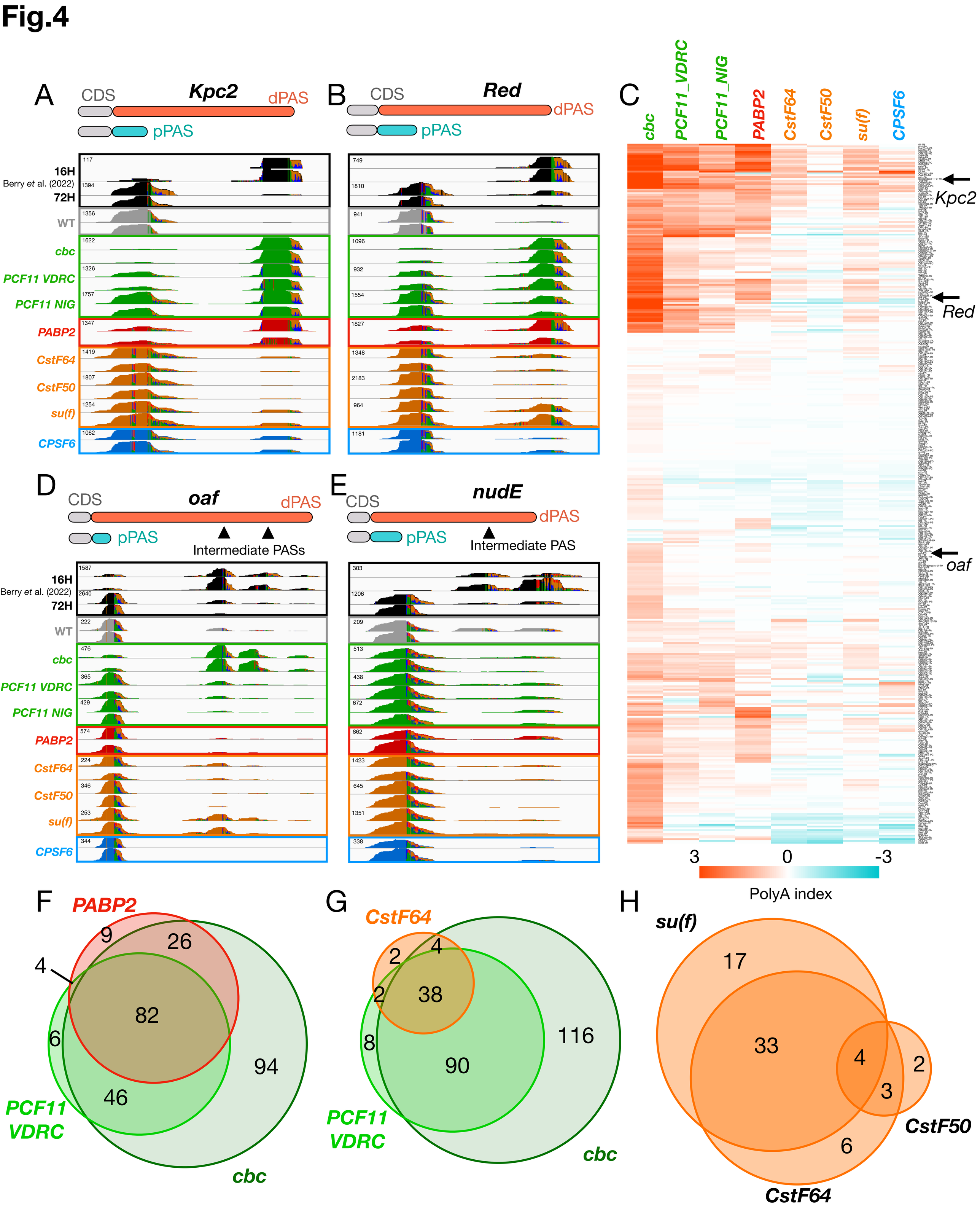

A
CDS
Kpc2
dPAS
pPAS
B
CDS
Red
dPAS
pPAS
C
cbc
PCF11_VDRC
PCF11_NIG
PABP2
CstF64
CstF50
su(f)
CPSF6
Kpc2
Red
oaf
3
0
-3
PolyA index
16H
Berry et al. (2022)
72H
WT
cbc
PCF11 VDRC
PCF11 NIG
PABP2
CstF64
CstF50
su(f)
CPSF6
D
CDS
oaf
dPAS
pPAS
Intermediate PASs
E
CDS
nudE
dPAS
pPAS
Intermediate PAS
F
PABP2
9
26
4
82
6
94
46
PCF11 VDRC
cbc
G
CstF64
2
4
2
38
8
116
90
PCF11 VDRC
cbc
H
su(f)
17
33
4
2
3
6
CstF50
CstF64

Fig.5

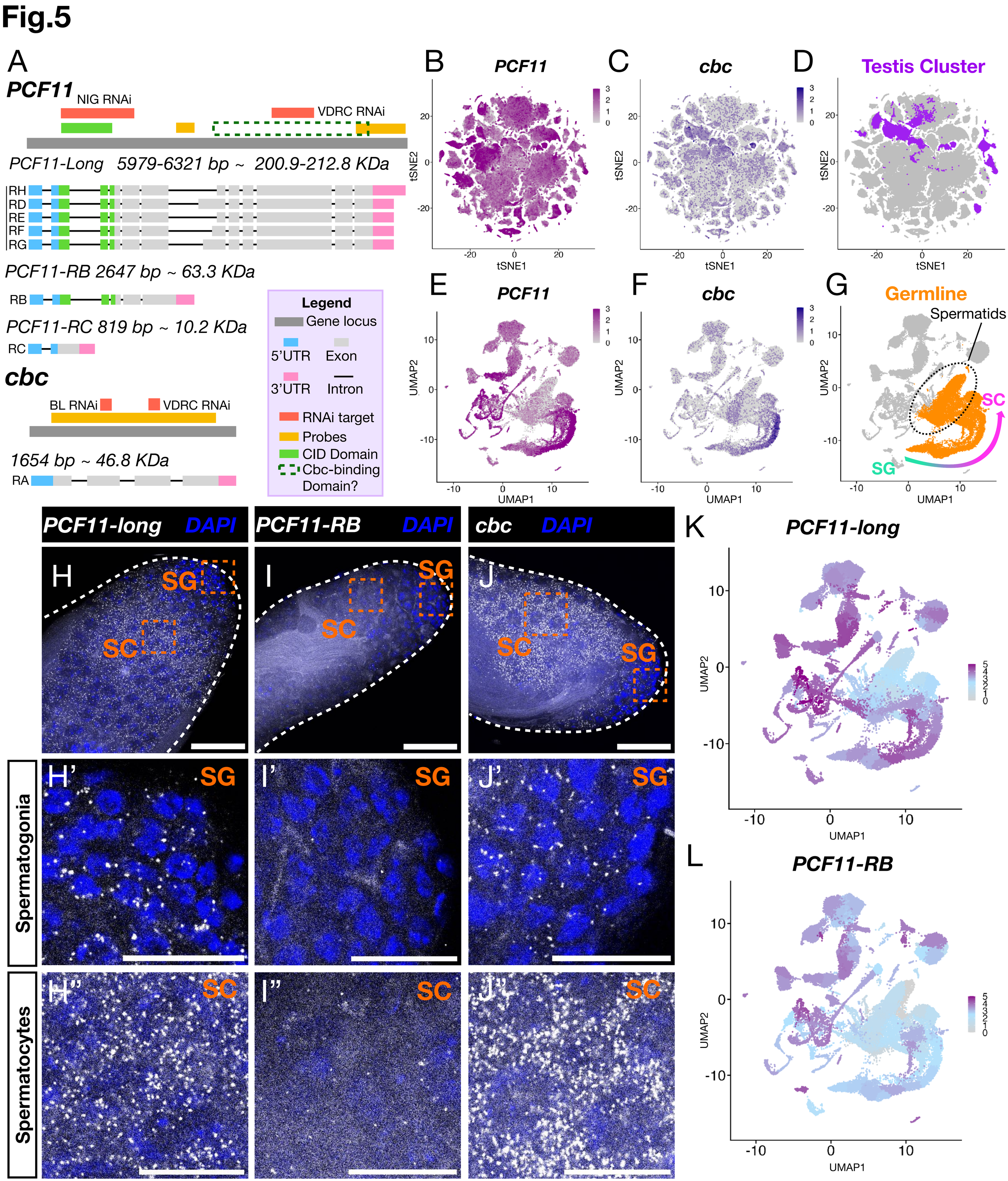

A
PCF11
NIG RNAi
VDRC RNAi
PCF11-Long 5979-6321 bp ~ 200.9-212.8 KDa
RH
RD
RE
RF
RG
PCF11-RB 2647 bp ~ 63.3 KDa
RB
PCF11-RC 819 bp ~ 10.2 KDa
RC
cbc
BL RNAi
VDRC RNAi
1654 bp ~ 46.8 KDa
RA
Legend
Gene locus
5'UTR
Exon
3'UTR
Intron
RNAi target
Probes
CID Domain
Cbc-binding Domain?
B
PCF11
C
cbc
D
Testis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

Fig.6

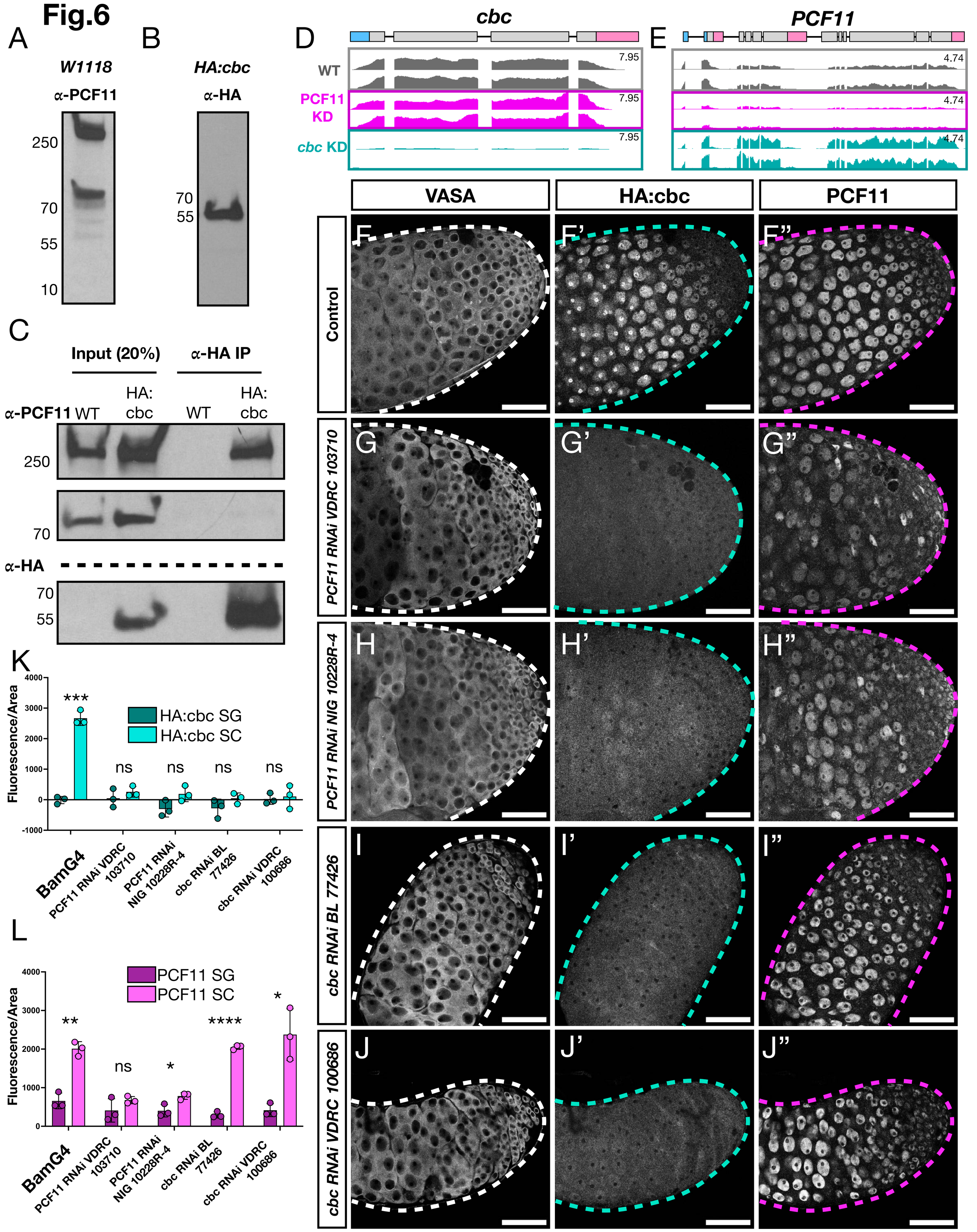
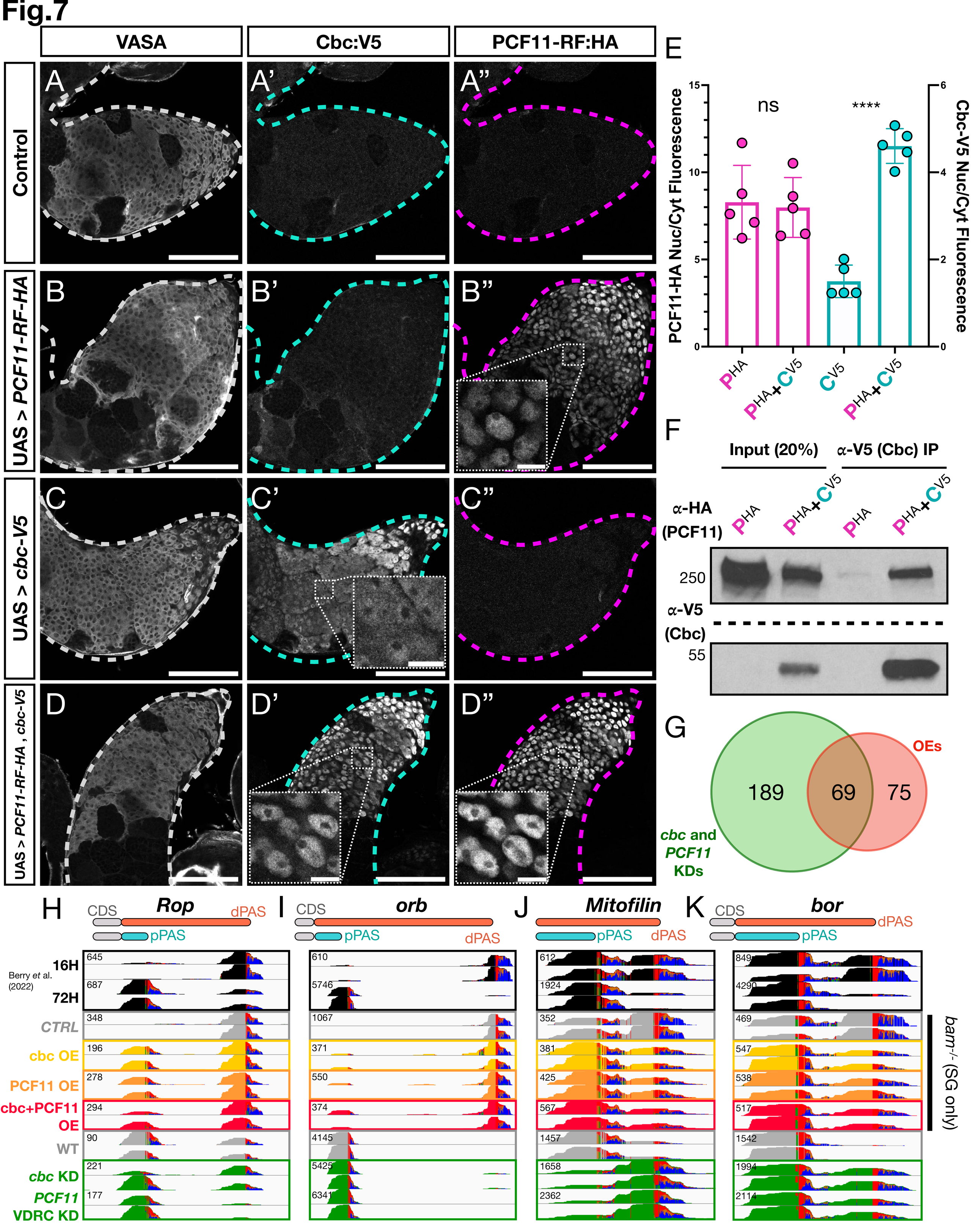

A
W1118
α-PCF11
250
70
55
10
B
HA:cbc
α-HA
70
55
C
Input (20%)
α-HA IP
α-PCF11
WT
HA: cbc
WT
HA: cbc
250
70
α-HA
70
55
D
cbc
WT
PCF11 KD
cbc KD
7.95
E
PCF11
4.74
VASA
HA:cbc
PCF11
Control
PCF11 RNAi VDRC 103710
PCF11 RNAi NIG 10228R-4
cbc RNAi BL 77426
cbc RNAi VDRC 100686
F
F'
F''
G
G'
G''
H
H'
H''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

Fig.7


VASA
Cbc:V5
PCF11-RF:HA
Control
UAS > PCF11-RF-HA
UAS > cbc-V5
UAS > PCF11-RF-HA , cbc-V5
A
A'
A''
B
B'
B''
C
C'
C''
D
D'
D''
E
ns
****
PCF11-HA Nuc/Cyt Fluorescence
Cbc-V5 Nuc/Cyt Fluorescence
PHA
PHA+CV5
CV5
PHA+CV5
F
Input (20%)
α-V5 (Cbc) IP
α-HA (PCF11)
250
α-V5 (Cbc)
55
G
OEs
189
69
75
cbc and PCF11 KDs
H
Rop
CDS
dPAS
pPAS
I
orb
J
Mitofilin
K
bor
Berry et al. (2022)
16H
72H
CTRL
cbc OE
PCF11 OE
cbc+PCF11 OE
WT
cbc KD
PCF11 VDRC KD
645
687
348
196
278
294
90
221
177
610
5746
1067
371
550
374
4145
5425
6341
612
1924
352
381
425
567
1457
1658
2362
849
4290
469
547
538
517
1542
1994
2114
bam-/- (SG only)