Title: I-KCKT allows dissection-free RNA profiling of adult *Drosophila* intestinal progenitor cells

Authors: Kasun Buddika*, Jingjing Xu*, Ishara S. Ariyapala, and Nicholas S. Sokol

* Co-first authors.

Affiliation: Department of Biology, Indiana University, Bloomington, IN 47405, USA

Correspondence: nsokol@indiana.edu (email)
812-856-6812 (phone)

Keywords: CLIP, eCLIP, PABP, FMRP, intestinal stem cell

Running Title: I-KCKT RNA profiling

23 **Summary Statement**

24 We report a dissection-free method to identify proximity-based RNA-protein
25 interactions in an *in vivo* stem cell population, enabling molecular analysis of these cells
26 at unprecedented speed and resolution.

27

Abstract

The adult Drosophila intestinal epithelium is a model system for stem cell biology, but its utility is limited by current biochemical methods that lack cell type resolution. Here, we describe a new proximity-based profiling method that relies upon a GAL4 driver, termed *intestinal-kickout-GAL4* (*I-KCKT-GAL4*), exclusively expressed in intestinal progenitor cells. This method used UV cross-linked whole animal frozen powder as its starting material to immunoprecipitate the RNA cargoes of transgenic epitope-tagged RNA binding proteins driven by *I-KCKT-GAL4*. When applied to the general mRNA-binder, poly(A)-binding protein, the RNA profile obtained by this method identified 98.8% of transcripts found after progenitor cell sorting, and had low background noise despite being derived from whole animal lysate. We also mapped the targets of the more selective RNA binder, Fragile Mental Retardation Protein, using enhanced CLIP, and report for the first time its binding motif in Drosophila cells. This method will therefore enable the RNA profiling of wildtype and mutant intestinal progenitor cells from intact flies exposed to normal and altered environments, as well as the identification of RNA-protein interactions critical for stem cell function.

Introduction

The adult *Drosophila* intestine has become a premier model for understanding the biology and behavior of resident stem cells in their native context (Li and Jasper, 2016). One key approach has been transcript profiling that characterizes the gene expression signatures of cell types in both homeostatic and perturbed conditions (see, for example, (Dutta et al., 2015; Hung et al., 2020)). This approach has relied on the manual dissection of intestines. Manual dissection is not only laborious and time-consuming, but also precludes certain types of analysis that require large amounts of intact, non-degraded starting material, such as immunoprecipitation (IP). Methods that rely upon the sequencing of immunoprecipitates associated with RNA-binding proteins (RBPs) to analyze post-transcriptional control mechanisms (e.g. RiboTag) are not currently feasible in adult Drosophila intestinal stem cells but have been critical for characterizing identity and differentiation mechanisms in other stem cell lineages (Baser et al., 2019; Sanz et al., 2009; Tahmasebi et al., 2018). There is therefore a need for improved methods to profile this stem cell population in Drosophila.

Manual dissection of intestines has been required because of the lack of tools to exclusively label and manipulate individual intestinal cell types. Such tools would enable the use of the whole fly as starting material, allowing for the rapid production of large amounts of starting material. GAL4 drivers are available that label the various individual cell types in the intestinal stem cell lineage, including the intestinal stem cells (ISCs) themselves as well as their transient progenitor cell daughters, termed enteroblasts (EB), and two differentiated cell types, enterocytes (ECs) and enteroendocrine cells (EEs) (Jiang and Edgar, 2009; Micchelli and Perrimon, 2006; Zeng et al., 2010). Such

67 GAL4 drivers can be used to express epitope-tagged transgenic proteins specifically in
68 these various intestinal cell types that can be immunopurified. However, these GAL4
69 drivers are either known or likely to be active elsewhere in the adult (Biteau et al.,
70 2010).

71 Here, we describe a method that uses a new intestinal progenitor GAL4 driver
72 that is not expressed outside the intestine to profile the general transcriptome as well as
73 specific RBP cargoes expressed in this cell type. We reasoned that an intersectional
74 approach could be used to design such a GAL4 driver. Intersectional methods limit
75 transgenic expression to cells in which two different enhancers are both active. In a
76 recombinase-mediated intersectional method, for example, the recombinase under the
77 control of one enhancer activates a GAL4 driver under the control of a second enhancer
78 via recombinase-mediated removal of an intervening stop cassette (Fig 1A). Such a
79 recombinase-based method involving a pan-intestinal enhancer and a progenitor
80 enhancer should limit expression to only intestinal progenitor cells. We designed our
81 transgenes with the KD recombinase (KDR), which was recently shown to mediate the
82 excision, or “kick-out”, of sequences between KDR target (KDRT) sites in *Drosophila*
83 cells (Nern et al., 2011). We chose to use KDR so that the resulting system could be
84 used in tandem with other recombinases, like FLP, for additional manipulation of
85 intestinal cells. Because this system is designed for intestine-specific activation of a
86 “kick-out” transgene, we referred to it as Intestinal-KiCK-ouT, or I-KCKT.
87

88 ## Results

89 ### *I-KCKT-GAL4* labels most intestinal progenitor cells

90 To build the transgenes needed for our system, we first searched for defined
91 pan-intestinal and progenitor enhancer sequences. Progenitor enhancer sequences
92 have previously been identified in the *miranda* (*mira*) gene locus (Bardin et al., 2010),
93 but the availability of a defined pan-intestinal enhancer was less clear. Transgenes with
94 regulatory sequences from some intestine-specific genes have been reported (e.g.
95 *mex1-GAL4*, *npc1b-GAL4*, etc. (Phillips and Thomas, 2006; Voght et al., 2007)), but
96 careful inspection found that these transgenes were either not expressed throughout the
97 intestine or in all cell types. We therefore took a candidate gene approach to identify
98 intestine-specific enhancers based on gene expression data reported in FlyAtlas 2
99 (Leader et al., 2018), testing DNA fragments from intestine-specific genes for intestinal
100 activity. KDR transgenes containing putative enhancer sequences from *βTrypsin* (*βTry*),
101 *θTrypsin* (*θTry*), *κTrypsin* (*κTry*), *CG10116*, or *CG18404* were tested for intestinal
102 activity in flies also harboring two additional transgenes, a *tubulin-KDRT-stop-KDRT-*
103 *GAL4.p65* "kickout" transgene and a *UAS-6XGFP* responder (Shearin et al., 2014).
104 GFP expression was monitored in both the larval and adult intestine for all resulting
105 strains except for *θTry*; in its case, only larvae were analyzed because adults of the
106 proper genotype failed to eclose. While all strains displayed some GFP expression in
107 the larval intestine, most GFP patterns were patchy and non-uniform (Fig 1B-G). In the
108 adult intestine, however, two KDR lines, *βTry-KDR* and *CG10116-KDR*, drove
109 expression throughout the tissue (Fig 1H, M). Careful inspection of these two indicated
110 that *CG10116-KDR* was expressed in most cells throughout the midgut, raising the

111 possibility that the associated enhancer fragment was pan-intestinal-specific and could
112 be used to exclusively label adult intestinal progenitor cells.
113 To test this possibility, we evaluated *CG10116-KDR* activity in the intestinal
114 progenitor cells of flies harboring a *mira*-containing stop cassette transgene as well as
115 *UAS-stinger-GFP*, a nuclear-localized GFP reporter (Barolo et al., 2000). For these
116 experiments, we analyzed strains containing two different *mira* transgenes, one
117 encoding an enhanced version of GAL4 with the p65 transcriptional activation domain
118 (*mira-KDRT-stop-KDRT-GAL4.p65*) and one with a non-modified version of GAL4 (*mira-*
119 *KDRT-stop-KDRT-GAL4*) that could be used in conjunction with the GAL80-dependent
120 temporal and regional gene expression targeting (TARGET) system for conditional
121 expression (McGuire et al., 2004). For brevity, we refer to the strains combining
122 *CG10116-KDR* with *mira-KDRT-stop-KDRT-GAL4.p65* or *mira-KDRT-stop-KDRT-GAL4*
123 as *I-KCKT-GAL4.p65* or *I-KCKT-GAL4*, respectively. To analyze *CG10116-KDR*-
124 mediated expression in these strains, we compared GFP expression with that of a
125 previously generated and validated progenitor reporter, *mira-His2A.mCherry.HA* (Miller
126 et al., 2020), in each of the five intestinal regions (Fig 2A-J). Quantification of the
127 percentage of mCherry+ cells that were also GFP+ indicated that ~100% of progenitor
128 cells were labeled in most regions of both *I-KCKT* strains (Fig 2P, Q). The only two
129 exceptions to this trend were regions 1 and 3 of *I-KCKT-GAL4.p65* intestines, where
130 only 58.7±15.5% (n=5) and 64.4±17.0% (n=5) of progenitor cells were labeled,
131 respectively. Importantly, we also found that no GFP+ cells were mCherry- in either
132 strain, indicating that no non-progenitor cells were labeled in *I-KCKT* strains. To confirm
133 this analysis, we also compared GFP expression driven by *I-KCKT-GAL4* to a second,

134 validated progenitor reporter, *esg-LacZ* (Micchelli and Perrimon, 2006), and found
135 similarly that almost all LacZ+ cells were also GFP+ and that no GFP+ cells were LacZ-
136 (Fig S1). The only exception was region 3, where ~69.9±9.0% of LacZ+ cells were
137 GFP+. This discrepancy may reflect that *esg-LacZ* labels a small subset of EEs in this
138 midgut area (Hung et al., 2020) Collectively, this analysis indicated that almost all
139 progenitor cells, but few if any other cells, were labeled in the intestines of *I-KCKT*
140 strains.
141
142 ***I-KCKT-GAL4* is not detected in non-intestinal tissue**
143 To determine whether intestinal progenitor cells were exclusively labeled in *I-*
144 *KCKT* adults, we performed two analyses. First, we crossed *I-KCKT* strains to *UAS-*
145 *6XGFP* and visually inspected adults for GFP expression. *I-KCKT* strains displayed
146 prominent 6XGFP fluorescence in the intestine but little, if any, elsewhere (Fig 3A, B).
147 For comparison, we similarly analyzed three other widely used drivers known to be
148 expressed in some or all intestinal progenitor cells: a *GawB* P-element insertion in the
149 *escargot* (*esg*) locus (*esg-GAL4*), a *GawB* P-element insertion in the *Delta* (*Dl*) locus
150 (*Dl-GAL4*), and a *Notch*-responsive GAL4 reporter that contains binding sites for the
151 Grainyhead and Suppressor of Hairless transcription factors (*gbe-GAL4*) (Micchelli and
152 Perrimon, 2006; Zeng et al., 2010). All three were detected in the intestine (Fig 3E-G);
153 *Dl-GAL4* and *gbe-GAL4* also displayed prominent non-intestinal expression while *esg-*
154 *GAL4* appeared more similar to the *I-KCKT* with regard to intestinal specificity. For a
155 more rigorous analysis, we also performed comparative Western Blot analysis for
156 6XGFP expression on three different protein extracts generated from each strain: whole

157 animal extract (total), extract from dissected gastrointestinal tracts that included the
158 malphigian tubules (intestine), and extract from all remaining non-intestinal tissue
159 (carcass) (Fig 3H). For both *I-KCKT* strains, GFP was detected in intestinal but not
160 carcass extract, and the amount in the intestine was roughly similar to the amount in the
161 total extract. In contrast, GFP was detected in both the intestine and carcass of the
162 three comparison strains, *esg-GAL4*, *Dl-GAL4*, and *gbe-GAL4*. We also tested the
163 conditional expression of *mira-KDRT-stop-KDRT-GAL4* by generating an *I-KCKT-*
164 *GAL4*^{TS} strain that harbored an ubiquitously expressed, temperature sensitive GAL80
165 transgene (*tub-GAL80*^{TS}) and then performing both analyses on flies that had been
166 incubated at the non-permissive (18°C) and permissive (30°C) temperatures (Fig 3H-J).
167 This analysis detected no GFP expression at the non-permissive temperature, and clear
168 expression at the permissive temperature. We noted that *I-KCKT-GAL4*^{TS}-driven GFP
169 signal can also be detected in the progenitor cells of the malphigian tubules
170 (arrowheads in Fig 3J), which are *esg*$^{+}$ cells that are related to midgut progenitor cells
171 (Singh et al., 2007). We also note that, like *esg-GAL4* but to a lesser degree, *I-KCKT-*
172 *GAL4*^{TS} lost its progenitor-specificity in the intestinal tissue of aged animals (Fig S2).
173 Collectively, these results indicated that, unlike other commonly used progenitor drivers,
174 *I-KCKT*-GAL4-based expression could not be detected outside the intestine.
175

176 ***ISC-KCKT-GAL4* specifically labels adult intestinal stem cells**

177 To further expand on the utility of the I-KCKT system, we investigated whether *I-*
178 *KCKT*-based GAL4 expression could be limited to either the ISC or EB subsets of
179 progenitor cells. To do so, we prepared strains in which the three *I-KCKT*^{TS} transgenes

180 were combined with a GAL4-silencing GAL80 transgene that contained either the EB-
181 specific *gbe* synthetic enhancer (*gbe-GAL80*) or a fragment from the *Delta* locus
182 reported to be active in ISCs (*GMR24H06-GAL80*) (Furriols and Bray, 2001; Guo et al.,
183 2013). We referred to these strains as *ISC-KCKT-GAL4*^{TS} and *EB-KCKT-GAL4*^{TS},
184 respectively. Like *I-KCKT-GAL4*^{TS}, both *ISC-KCKT-GAL4*^{TS} and *EB-KCKT-GAL4*^{TS}
185 activity were intestine-specific based on visual and/or Western blot analysis (Fig 3C, D,
186 H). To evaluate the cell type specificity of this activity, we drove GFP expression in both
187 strains and compared it to a dual reporter combination that effectively distinguishes
188 ISCs and EBs (Fig S3). These dual reporters are the progenitor marker *mira-*
189 *His2A.mCherry.HA* (Miller et al., 2020) and the EB-specific marker *3Xgbe-*
190 *smGFP.V5.nls* (Buddika et al., 2020a). For this analysis, we scored mCherry+, V5- cells
191 as ISCs and mCherry+, V5+ cells as EBs. Most ISCs were specifically labeled in the
192 *ISC-KCKT-GAL4*^{TS} strain: 84.1±12.3% (n=5) of ISCs were labeled whereas 11.2±10.0%
193 (n=5) of EBs were labeled across all five regions of the intestine (Fig 2K-O, R, S). In
194 contrast, EB-specific labeling in *EB-KCKT-GAL4*^{TS} was less effective: only 30.7±20.1%
195 (n=5) of EBs were labeled whereas 57±19.8% (n=5) of ISCs were labeled across all five
196 regions of the intestine (Fig S4). Various possibilities could explain this *EB-KCKT-*
197 *GAL4*^{TS} result, including that the GMR24H06 enhancer fragment was not active in most
198 ISCs and/or that GAL80 activity perdured in EB cells. Nevertheless, this analysis
199 indicated that *ISC-KCKT-GAL4*^{TS} specifically labeled most ISCs.
200
201 **CLIP-seq of PABP using *I-KCKT-GAL4* identifies progenitor expressed genes**

202 We next investigated whether these *I-KCKT* strains could be used to streamline
203 current methods to molecularly profile progenitor cells. RNA-seq profiling methods that
204 rely on *esg-GAL4*, for example, require labor-intensive dissection of hundreds of
205 intestines followed by digestion of this tissue to release labeled progenitor cells for
206 fluorescent activated cell sorting (FACS) (Dutta et al., 2015; Fast et al., 2020; Korzelius
207 et al., 2019; Li et al., 2018) (Fig 4A, left). Because *I-KCKT-GAL4* activity was limited to
208 intestinal progenitor cells, we tested whether a cross-linking immunoprecipitation and
209 sequencing (CLIP-seq) analysis of *I-KCKT-GAL4*-driven, FLAG-tagged poly-A Binding
210 Protein (PABP) that used whole animal lysate as its starting material would recover
211 progenitor cell RNA (Fig 4A, right). PABP is a general mRNA-binding protein that has
212 been used to profile the mRNA transcriptomes of other cell types (Hwang et al., 2016;
213 Yang et al., 2005). We first verified that the FLAG-tagged version of PABP that had
214 previously been used to profile mRNAs in *Drosophila* photoreceptor cells was
215 expressed in progenitor cells when crossed to *I-KCKT-GAL4*^{TS} (Fig S5) (Yang et al.,
216 2005). Then, we crushed frozen flies from this strain to generate a fine powder,
217 subjected this powder to UV crosslinking to covalently link protein and RNA complexes,
218 immunoprecipitated PABP using anti-FLAG beads, and recovered the associated
219 RNAs. Independent libraries were prepared from duplicate PABP immunoprecipitates
220 (PABP IPs). In parallel, two libraries were also prepared from RNA extracted from the
221 starting whole animal lysate, and we used these “input” libraries as our normalization
222 controls. Differential gene expression analysis found 1,661 transcripts enriched in the
223 PABP IP samples and 3,293 enriched in input (Fig. 4B, Table S1). Representative

224 examples of these respective classes included the progenitor-enriched *esg* transcript
225 and the ovary-enriched *oskar* transcript (*osk*) (Fig. 4C-D).
226 To evaluate whether the PABP IP-enriched transcripts included transcripts
227 known to be enriched in progenitor cells, it was cross-referenced to two published
228 datasets. First, we compared our results to transcriptome profiles obtained by Doupé,
229 who used a DamID based method to identify progenitor-enriched and EC-enriched
230 transcripts (Doupe et al., 2018). Twenty percent of the 1,661 were present in Doupé's
231 progenitor-enriched transcripts, whereas only 4% were present in the EC-enriched set
232 (Fig. 4E-F). In addition, we re-analyzed published RNA-seq datasets from FACS-
233 isolated progenitors, ECs, and EEs (Dutta et al., 2015), identified transcripts enriched in
234 progenitor cells when compared to either ECs or EEs and found that 31%-32% of these
235 enriched transcripts overlapped with transcripts enriched in the PABP IP (Fig. 4G-H,
236 Table S2). To benchmark these comparisons with the Doupé and Dutta datasets, we
237 performed analogous comparisons between them and an RNA-seq dataset of *esg*+
238 progenitor cells that we generated using FACS and the same library preparation kit as
239 for the PABP IP libraries (Buddika et al., 2020b), and found similar levels of overlap
240 (Fig. S6A-D). This analysis indicated, therefore, that the transcripts most highly
241 represented in the PABP IP relative to whole animal lysate included a substantial
242 number of genes previously shown to be enriched in progenitor cells.
243 We next focused on comparing the entire PABP-associated transcriptome to the
244 RNA-seq profile of FAC sorted progenitor cells, using four previously reported RNA-seq
245 datasets. These datasets were reported in (Buddika et al., 2020b; Dutta et al., 2015;
246 Fast et al., 2020; Korzelius et al., 2019), and we refer to them as Buddika, Datta, Fast,

247 and Korzelius, respectively. Read counts for the four FACS datasets as well as the
248 PABP IP dataset were normalized in parallel to facilitate direct comparison; a transcript
249 was considered to be expressed in a dataset if it had a normalized expression value
250 greater than 10 in each replicate. Only 61 genes found in all four FACS datasets were
251 not present in the PABP IP dataset, indicating that the PABP IP identified 98.8% (4,966
252 out of 5,027) of core progenitor transcripts common to all four distinct FACS datasets
253 (Fig 4I). An additional 1,839 transcripts were found in the PABP IP, and 76.2% of these
254 were also found in at least one other FACS dataset. Four hundred and thirty-seven
255 transcripts (6.4%) were unique to the PABP IP, a number in line with the numbers
256 unique to each FACS dataset and that likely reflected differences caused by the
257 technical approach. These could include normal deviations associated with distinct
258 experimental settings, as well as differences related to the preparation of flash frozen
259 versus mechanically disrupted samples. Altogether, these results indicated that the
260 PABP IP method effectively identified progenitor transcripts.

261 To evaluate the significance of the overlap between these five datasets,
262 Spearman correlation matrices were generated and visualized as correlograms. As a
263 negative control for this analysis, we also prepared a sixth dataset of head tissue data
264 from FlyAtlas 2 (Leader et al., 2018). As expected, the PABP IP correlated well with all
265 FACS based progenitor datasets while, furthermore, all five of these datasets showed a
266 similar, lower correlation with the non-intestinal dataset (Fig. 4J). We then used the
267 GeneOverlap R package to perform Fisher's exact test to evaluate the statistical
268 significance of the overlap between different datasets. The Fisher's exact test also
269 computes both an odds ratio and a Jaccard index, which represents the strength of

270 association and similarity between two datasets, respectively. The PABP IP scored
271 highly in both these indices when compared to each of the four FAC-sorted datasets
272 and significantly higher than either the total input or head tissue datasets, further
273 indicating the similarity between the PABP IP and FACS-based datasets (Fig. 4K).
274 Finally, we evaluated whether the PABP IP contained progenitor-enriched
275 transcripts that we identified from systematic re-analysis of the four FACS-based
276 datasets. For this re-analysis, differential expression analysis was used to select for
277 transcripts enriched in Buddika, Datta, Fast, and Korzelius compared to two sources of
278 whole animal data, total input from the PABP experiment or reanalyzed whole female
279 adult dataset from FlyAtlas 2. The PABP IP was then compared to each of these
280 computed progenitor enriched datasets. As expected, PABP IP enriched genes showed
281 significant low p-values and high odds ratio and Jaccard index scores, strengthening the
282 positive relationship of PABP and FACS-based datasets (Fig. S6E). Altogether, this
283 bioinformatic analysis supported the conclusion that the PABP IP dataset represented
284 the gene expression profiles of progenitor cells.
285
286 **CLIP-seq of I-KCKT-GAL4-driven PABP has little background**
287 Having established that the PABP IP dataset contained progenitor genes, we
288 also sought to evaluate how much non-specific background was represented in this
289 dataset. Such noise might come from the post-lysis association of transgenic PABP with
290 mRNAs from non-progenitor cells (Mili and Steitz, 2004). Hypothesizing that the most
291 abundant transcripts in the whole animal lysate might be most prone to associate with
292 PABP post-lysis, we first identified the top 10% of genes in the input based on the

293 normalized expression values and found 1049 genes. Only 207 of these genes were
294 also identified in the PABP IP enriched gene set, suggesting that transcripts abundantly
295 expressed in the starting lysate were not non-specifically recovered in the IP (Fig. 4L).
296 Supporting this observation, we identified 854 genes in the input that were expected not
297 to be present in the PABP IP (Fig. S6F) and confirmed that only 4 of these transcripts
298 were present in the PABP IP enriched gene list (Fig. 4M). Finally, we again used
299 reanalyzed FlyAtlas 2 data to identify genes enriched in fat body, head, ovary and testis
300 relative to midgut, and then compared those tissue-enriched genes to the genes either
301 enriched in or depleted from the PABP IP. We found that odds ratio and Jaccard index
302 scores were low for PABP-enriched genes, indicating minimal overlap, and significantly
303 higher for PABP-depleted genes (Fig. 4N). As a positive control for this analysis, we
304 performed the same comparisons with genes enriched in the Fly Atlas 2 midgut dataset
305 as well as genes enriched in the Buddika FACS dataset. As expected, PABP IP
306 enriched genes show highest association with genes enriched in FACS-isolated
307 progenitors, and significant overlap was also observed with the midgut enriched gene
308 list (Fig. 4N). Collectively, this analysis indicated that the PABP dataset contained
309 minimal background noise from other tissues, likely due to the stringency of the washes
310 after the crosslinking step.
311

312 ***I-KCKT*-based eCLIP analysis identifies FMRP-bound mRNAs in the intestine**

313 To illustrate the breadth of its possible applications, we employed *I-KCKT-GAL4*
314 to identify the RNA cargo of a more selective RBP, Fragile X Mental Retardation Protein
315 (FMRP), specifically in progenitor cells using enhanced Crosslinking

316 Immunoprecipitation sequencing (eCLIP-seq). FMRP limits ISC expansion but its mRNA
317 targets in these cells are unknown (Luhur et al., 2017). The eCLIP method contains a
318 number of key modifications in comparison to other CLIP methods, including adapter
319 ligation steps that map the exact position of crosslinking and the preparation of a size-
320 matched input (SMI) control sample for stringent normalization (Van Nostrand et al.,
321 2016). We first verified that *UAS-FMRP.FLAG.GFP* in combination with *I-KCKT-GAL4*^{TS}
322 was detected in progenitor cells after two and ten days of induction and that it caused a
323 reduction in progenitor cell number when induced for ten but not two days (Fig S5). We
324 then prepared FMRP IP and SMI libraries from whole animals collected at the two-day
325 timepoint (Fig S7A-B). After removing rRNA reads, multimapping reads, and duplicated
326 reads, about 8% (8,438,543), 5% (2,487,118) and 7% (2,545,165) of total reads were
327 recovered as usable reads from the SMI and two IP samples, UV1 and UV2,
328 respectively (Fig S7C), consistent with the recovery rate from other eCLIP analyses
329 (Van Nostrand et al., 2016).
330 Peak calling analysis identified 13,297 reproducible FMRP binding sites across
331 1,829 genes (Table S3) and indicated that there was high correlation between replicates
332 (Fig S7D), demonstrating the robustness of our modified eCLIP method. Analysis of
333 peak distribution indicated that >85% of FMRP binding sites were in protein-coding
334 transcripts and, within these transcripts, the majority of sites were in coding sequences
335 (Fig 5A, B). Both of these results were consistent with previous characterizations of
336 FMRP distribution in multiple species, including from proximity-based analyses in mice
337 and human tissues as well as activity-based analysis in fly cells (Darnell et al., 2011; Li
338 et al., 2020; McMahon et al., 2016). In addition, we noted that this analysis identified

339 several confirmed direct targets of Drosophila FMRP found by genetic or targeted
340 biochemical means, including *CaMKII*, *chic*, *Dscam1, futsch, ninaE*, *rg* and *tral*
341 (Cvetkovska et al., 2013; Kim et al., 2013; Monzo et al., 2010; Reeve et al., 2005; Sears
342 et al., 2019; Sudhakaran et al., 2014; Wang et al., 2017; Zhang et al., 2001). As the first
343 CLIP-based assay to characterize Drosophila FMRP, this analysis also identified
344 CAUUG(A/U) as its top binding motif (Fig 5C), consistent with prior identification of the
345 AUUG sequence in one of the top binding motifs of human FXR1 (Feng et al., 2019).
346 The FMRP mRNA cargo includes ~20% of the protein-coding transcripts
347 expressed in progenitor cells that were identified by the PABP IP analysis described
348 above (Fig 5D). These FMRP targets were significantly associated with processes
349 related to stem cell proliferation, stem cell maintenance, cell differentiation, and
350 translation repressor activity (Fig. S7E), consistent with the known roles of FMRP in
351 stem cell populations such as ISCs (Luhur et al., 2017). One representative example of
352 these FMRP targets was *esg* (Fig 5E), which plays critical roles in progenitor cells
353 where it is a known target of post-transcriptional control (Antonello et al., 2015;
354 Korzelius et al., 2014). We also noted that ~12% of the FMRP cargo are from genes not
355 identified in either PABP IP or progenitor cell RNA-seq (Fig 5D, S7F). This signal could
356 therefore represent post-lysis association of FMRP.FLAG.GFP with non-progenitor
357 mRNAs. However, none of 186 FMRP target genes that were not represented in the
358 PABP IP were among the 1049 most abundant transcripts present in the whole fly input
359 (Fig S7G). We therefore favor the hypothesis that these FMRP targets represent
360 weakly-expressed progenitor transcripts that were not recovered in the PABP IP
361 because of their low level of expression.

362 To verify these eCLIP-identified FMRP targets, we first performed qPCR on RNA
363 immunoprecipitated with endogenous FMRP from wildtype intestines (Fig S7H). We
364 chose seven target genes (*apt*, *Egfr*, *esg*, *Mes2*, *shg*, *Tet*, and *Unr*) as well as five
365 negative controls (*CG15784*, *CG5767*, *gapdh1*, *gapdh2*, and *Pyk*) for this analysis. As
366 expected, the FMRP target genes showed between 2- and 18-fold enrichment in FMRP
367 IP relative to whole intestine input, while the negative controls were either not
368 significantly enriched or significantly de-enriched in the FMRP IP (Fig 5F). In addition,
369 we found that the length distribution of the FMRP-bound mRNAs was significantly
370 longer than the other protein-coding transcripts expressed in progenitor cells (Fig 5G),
371 consistent with recent ribosome foot printing that showed that FMRP preferentially
372 regulates the translation of large proteins in Drosophila oocytes (Greenblatt and
373 Spradling, 2018). Altogether, these results indicated that applying eCLIP to whole
374 animal tissue that contained progenitor cells that expressed FLAG-tagged FMRP
375 effectively recovered the FMRP cargo from these cells.

376 ## Discussion

377 Here we describe a method to purify the RNA cargoes associated with intestinal
378 progenitor RBPs using whole animal lysate as the starting material. Because it is GAL4-
379 based, this method will have wide applicability for profiling the general PABP-bound
380 transcriptome of wildtype and, when coupled with UAS-RNAi transgenes, mutant
381 progenitor cells. In addition, it opens the possibility of molecularly characterizing the
382 stem cell targets of selective RBPs, a growing number of which have been shown to
383 play critical roles in progenitor cells (e.g. LIN28, FMRP, SPEN, TIS11 (Andriatsilavo et
384 al., 2018; Chen et al., 2015; Luhur et al., 2017; McClelland et al., 2017)). Finally, this
385 method can be used in conjunction with additional GAL4-based mRNA profiling
386 methods, such as TU-tagging and ribosome profiling (Chen and Dickman, 2017; Hida et
387 al., 2017), as well as methods that target other classes of RNAs, including microRNAs
388 (Luhur et al., 2013). While the method is currently limited to progenitor cells, its
389 applicability could be further expanded by generating kickout transgenes that drive
390 expression in additional intestinal cell types such as EEs and ECs. Furthermore, while
391 we focus on protein-RNA interactions in this study, we also expect that *I-KCKT-GAL4* in
392 combination with mass-spectrometry can be used to identify protein-protein interactions
393 and thereby characterize protein complexes in progenitor cells.

394 The I-KCKT method offers a number of advantages over current progenitor cell
395 profiling methods, which typically involve the FAC sorting of progenitor cells from
396 dissected and then mechanically disrupted intestines. The use of whole animals rather
397 than dissected intestines as starting material greatly expedites sample preparation,
398 enabling the analysis of larger numbers of samples that could be used to test additional

399 conditions and manipulations. In addition, because sample preparation is fast, this
400 approach may effectively capture labile RNA profiles. Along these lines, the elimination
401 of the FAC sorting step will also likely improve the accuracy of results, since FAC
402 sorting is known to affect gene expression profiles likely due to the time and mechanical
403 disruption associated with this step (Andra et al., 2020; Richardson et al., 2015). In
404 addition, this method could make molecular profiling of progenitors more accessible,
405 since cell sorting runs can be costly and the equipment is not always available.

406 While the conditional expression of transgenic PABP to retrieve bulk mRNA has
407 been used to approximate the transcriptomes of specific cell types in a variety of
408 species (Roy et al., 2002; Tenenbaum et al., 2000; Yang et al., 2005), some issues
409 regarding PABP should be considered when using this approach. PABP displays some
410 differential binding affinity to poly(A) RNAs that can introduce bias (Yang et al., 2005). In
411 addition, retrieval of RNAs with poly(A) tails misses deadenylated mRNAs as well as
412 some noncoding RNAs, thereby biasing resulting datasets towards actively translating
413 coding transcripts. Finally, the cell toxicity associated with transgenic expression of
414 PABP likely alters endogenous gene expression. Nevertheless, the cargo of PABP is
415 better reflective of cell transcriptomes than the cargos of other general RNA binders
416 and, furthermore, can be analyzed with paperCLIP to finely map 3'UTR ends and
417 uncover alternative polyadenylation sites in a cell type specific manner (Hwang et al.,
418 2016; Tenenbaum et al., 2000). It should also be noted that PABP IP results are well
419 suited for comparison with RNA-seq libraries that are generated via poly(A)-selection,
420 rather than rRNA depletion, since both methods select for polyadenylated transcripts.

421 As the first CLIP-based analysis of Drosophila FMRP, this study reports both
422 novel mRNA targets as well as the binding motif of this protein in intestinal progenitor
423 cells. The set of FMRP targets identified here displays partial overlap with those
424 identified in two recent studies that used either ribosome profiling or a proximity-based
425 activity assay to identify FMRP targets in oocytes, cultured cells, or neurons (Greenblatt
426 and Spradling, 2018; McMahon et al., 2016). Differences in the targets reported by
427 these studies likely reflect not only their distinct technical approaches but also the
428 variety of cell types analyzed, since FMRP is known to display cell type binding
429 preferences in other species (Maurin et al., 2018). Because loss of FMRP leads to
430 Fragile X Syndrome (FXS), the leading form of human intellectual disability in humans,
431 analysis of FMRP has focused on its activity in differentiated neurons. However, FMRP
432 also functions in stem cell populations, and its dysregulation in these cells likely
433 contribute to understudied FXS symptoms including elevated brain size, accelerated
434 growth, and gastrointestinal problems (Luhur et al., 2017). Thus, future analysis of the
435 stem cell targets of FMRP identified here may characterize regulatory relationships that
436 are relevant to therapies designed to treat the entire repertoire of FXS symptoms.
437 A current key limitation of the I-KCKT method is its purification of epitope-tagged
438 UAS-based transgenic protein rather than endogenous protein. This raises the concern
439 that the transgenic protein might be expressed at higher-than-endogenous levels, which
440 might cause inappropriate interactions with non-target mRNAs. Because the TARGET
441 system controlling transgene expression is temperature sensitive, this concern can be
442 addressed by testing multiple temperatures in the permissive range to identify
443 conditions supporting physiological protein expression. As an alternative approach to

444 address this concern, our current effort is focused on modifying endogenous RBP loci to
445 contain FRT-flanked epitope tags. I-KCKT-based expression of FLP should then lead to
446 the production of epitope-tagged endogenous protein in progenitor cells that would be
447 available for eCLIP-based analysis. Thus, we expect that I-KCKT-based methods will
448 allow unprecedented analysis of RNA-based mechanisms of progenitor cells.
449

450 ## Materials and Methods

451

452 ***Drosophila strains and husbandry***

453 All fly strains were cultured on standard Bloomington *Drosophila* stock center
454 media (https://bdsc.indiana.edu/information/recipes/bloomfood.html). Flies were reared
455 in 18°C, 25°C and 29°C incubators set for a 12hr light/dark schedule and 65% humidity.
456 The genotypes of all strains used in this study are listed in Table S4. Transgenesis to
457 create new strains was performed by Rainbow Transgenic Flies using plasmid DNA
458 described below. Additional stocks were obtained from the Bloomington Drosophila
459 Stock Center (*{20XUAS-6XGFP}attP2*, *P{UAS-Stinger}2*, *P{tubP-GAL80[ts]}20*, *P{UAS-*
460 *dPF}D*), Steven Hou (*P{Su(H)GBE-GAL4*, *P{GawB}Dl*$^{05151-G}$), the Kyoto Drosophila
461 Stock Center (*P{GawB}NP5130*), and Steve Stowers (*{20XUAS-DSCP-6XGFP}attP2*).
462 For GAL80-dependent conditional expression experiments, flies were reared at 18°C,
463 collected over 2 days, and then shifted to 30°C for up to 10 days before analysis. The
464 following strains have been deposited at the Bloomington Drosophila Stock Center: (**i**) *I-*
465 *KCKT-GAL4.p65* (BDSC #. Full genotype: *mira-KDRT>-dSTOP-KDRT>-*
466 *GAL4.p65}attP40* ; *{CG10116-KD.PEST}attP2*), (**ii**) *I-KCKT-GAL4*^{TS} (BDSC #. Full
467 genotype: *{mira -KDRT>-dSTOP-KDRT>-GAL4}attP40, P{tubP-GAL80[ts]}20 ;*
468 *{CG10116-KD.PEST}attP2*), (**iii**) *ISC- KCKT-GAL4*^{TS} (BDSC #. Full genotype*: {gbe-*
469 *GAL80}ZH-2A ; {mira -KDRT>-dSTOP-KDRT>-GAL4}attP40, P{tubP-GAL80[ts]}20 ;*
470 *{CG10116-KD.PEST}attP2*).

471

472 ***Transgenes***

473 *KD enhancer transgenes*: PCR products containing enhancer fragments amplified from
474 genomic DNA were cloned into the HindIII and AatII sites of *pJFRC161* (*20XUAS-IVS-*
475 *KD.PEST*, a gift from Gerald Rubin, Addgene plasmid 32140) using HiFi DNA Assembly
476 Mix (New England Biolabs). Oligo pairs used to amplify enhancer fragments from *βTry*
477 (3437/3438), *θTry* (3439/3440), *κTry* (3446/3447), *CG18404* (3443/3444 for the 2.6 kb
478 fragment, 3445/3443 for the 0.5kb fragment), and *CG10116* (3441/3442) are shown in
479 parentheses, and oligo sequences are reported in Table S5. Junctions of resulting
480 plasmids were verified by sequencing prior to the preparation of plasmid DNA for
481 transgenesis. *KD* enhancer transgenes were inserted into the *attP2* landing site.
482
483 *KDRT GAL4 transgenes*: *pJFRC164* (*21XUAS-KDRT>-dSTOP-KDRT>-myr.RFP*, a gift
484 from Gerald Rubin, Addgene plasmid 32141) was used as a backbone for KDRT-
485 flanked stop cassette plasmids. Four-way HiFi DNA Assembly reactions were
486 performed with (i) 7.2kb HindIII/XbaI-digested *pJFRC164* plasmid backbone, (ii) 1.4kb
487 AatII/XhoI-digested *pJFRC164* KDRT>-dSTOP-KDRT fragment, (iii) PCR-amplified
488 GAL4-containing fragments, and (iv) PCR amplified enhancer-containing fragments.
489 GAL4.p65 was amplified from *pBPGAL4.2.p65Uw* (a gift from Gerald Rubin, Addgene
490 plasmid 26229) with oligo pair 3433/3434, while GAL4 sequence was amplified from
491 *pBPGAL4.1Uw* (a gift from Gerald Rubin, Addgene plasmid # 26226) with oligo pair
492 4401/4402. Enhancer fragments include: (i) 2.6 kb *tubulin* fragment amplified with oligo
493 pair 3626/3627, (ii) 2.6 kb 5' and 1.6 kb 3' *miranda* fragments amplified with oligo pairs
494 3489/3490 and 3491/3492, respectively, and (iii) 3.4 kb *delta* fragment amplified with
495 oligo pair 3431/3432. Plasmid junctions and open reading frames were verified by

496 sequencing prior to the preparation of DNA for transgenesis. Oligo sequences are
497 reported in Table S5, and complete plasmid sequences are available upon request.
498 KDRT transgenes were inserted into the *attP40* landing site.
499
500 *GAL80 transgenes*: For *GMR24H06-GAL80*, the R24H06 enhancer fragment was PCR
501 amplified from genomic DNA and subcloned upstream of the GAL80 open reading
502 frame in a *pATTB*-containing transformation plasmid. *3xgbe-GAL80* was generated in a
503 similar manner except using oligos encoding three tandem copies of *gbe* sequence
504 (Furriols and Bray, 2001). Both of these *GAL80* transgenes were inserted into the *ZH-*
505 *2A* landing site.
506
507 ***Dissections and immunostaining***
508 Adult female flies were dissected in ice cold 1xPBS and fixed in 4%
509 paraformaldehyde (Electron Microscopy Sciences) in PBS for 45 min. For all samples to
510 be stained with antibodies other than anti-Dl, tissue was then washed in 1xPBT (1xPBS,
511 0.1% Triton X-100) and then blocked (1xPBT, 0.5% Bovine Serum Albumin) for at least
512 45 min. Samples were incubated at 4°C overnight with primary antibodies, including
513 rabbit anti-RFP (Clontech, 1:1000), mouse anti-V5 (MCA1360GA, Bio-Rad, 1:250),
514 chicken anti-LacZ (ab9361, Abcam, 1:2000), mouse anti-FLAG (F1804, Sigma, 1:1000),
515 and rabbit anti-HA (3724S, Cell Signaling Technology, 1:1000). The following day,
516 samples were washed in 1xPBT and incubated for 2-3 hours with secondary antibodies,
517 including AlexaFluor-488 and -568-conjugated goat anti-rabbit, -mouse, -rat and -
518 chicken antibodies (Life Technologies, 1:1000). Finally, samples were washed in

519 1xPBT, including one wash with DAPI (1:1000), and mounted in Vectashield mounting
520 medium (Vector Laboratories).
521 For staining with anti-Dl antibodies, an alternative sample preparation scheme
522 adapted from (Lin et al., 2008) was followed. Briefly, intestines were dissected in ice-
523 cold Grace's Insect Medium (Lonza Bioscience) and fixed in a 1:1 (V/V) mixture of
524 heptane (Sigma) and 4% paraformaldehyde (Electron Microscopy Sciences) in water for
525 15 min. The bottom aqueous paraformaldehyde layer was removed, 500µl of ice-cold
526 methanol was added, and shook vigorously for 30 seconds. The methanol/heptane
527 mixture was removed and incubated with 1ml ice-cold methanol for 5 min. Next,
528 samples were gradually rehydrated with a series of 0.3% PBT (1xPBS, 0.3% Triton X-
529 100):Methanol (3:7, 1:1, 7:3) washes, washed with 0.3% PBT alone for another 5 min
530 and blocked (0.3% PBT, 0.5% Bovine Serum Albumin) for at least 45 min. The primary
531 antibody was mouse anti-Delta (C594.9B, Developmental Studies Hybridoma Bank
532 [DSHB], 1:500), and secondary antibodies are described above. Samples were
533 mounted in ProLong Diamond mounting medium (Invitrogen).
534
535 ***Microscopy and image processing***
536 Images of whole flies were collected on a Zeiss Axio Zoom microscope and
537 images of dissected intestines were collected on a Leica SP8 Scanning Confocal
538 microscope. Samples to be compared were collected under identical settings on the
539 same day, image files were adjusted simultaneously using Adobe Photoshop CC, and
540 figures were assembled using Adobe Illustrator CC.
541

542 ***Western blot analysis***

543 Female *Drosophila* flies were used for protein isolation. Extract was prepared
544 from either whole animals or separately from the intestines and remaining carcass
545 tissue of dissected animals. Tissues were lysed in protein lysis buffer (25mM Tris-HCl
546 pH 7.5, 150mM KCl, 5mM $MgCl_2$, 1% NP-40, 0.5mM DTT, 1x Complete protease
547 inhibitor cocktail (Sigma)), protein extracts were resolved on a 4-20% gradient
548 polyacrylamide gel (Bio-Rad), transferred to Immobilon-P membrane (Millipore) and
549 probed with rabbit anti-GFP (ab290, Abcam, 1:10,000) or mouse anti-a-tubulin (12G10,
550 Developmental Studies Hybridoma Bank, 1:1000) antibodies. For IP verification, blots
551 were stained with mouse anti-FLAG (F1804, Sigma, 1:3500) or mouse anti-FMRP
552 (5A11, Developmental Studies Hybridoma Bank, 1:500). Subsequently, blots were
553 washed extensively with 1xTBST (1xTBS, 0.1% Tween-20) and incubated with anti-
554 rabbit or -mouse conjugated HRP secondary antibodies. After extensive secondary
555 washes with 1xTBST, blots were treated with ECL-detection reagents (Thermo
556 Scientific) and finally exposed to chemiluminescence films (GE Healthcare).

557

558 ***RNA immunoprecipitation, library construction, and sequencing***

559 Whole flies were collected and immediately snap frozen in liquid nitrogen and stored at -
560 80°C until sample preparation. Two hundred flies (100 male and 100 female) of each
561 sample were ground into a fine powder using a pre-cooled mortar and pestle on dry ice.
562 The powder was irradiated three times at 120 mJ per cm2 in a UV cross-linker
563 (Stratagene), with mixing between each irradiation cycle to maximize surface exposure.
564 The fly powder was transferred to a 2ml tube containing 1ml lysis buffer (150mM NaCl,

565 50mM Tris-HCl, pH7.5, 1mM EDTA, 1% Triton X-100, 1% Sodium Deoxycholate, 0.1%
566 SDS, 5X Complete Protease Inhibitor Cocktail (Sigma, 11836145001), 4U/ml RNase
567 Inhibitor Murine (BioLabs, M0314L) using a cold spatula, and the remaining fly powder
568 was washed off the spatula with an additional 1ml lysis buffer. Tubes were incubated on
569 ice for 15-30 min, with mixing every 5 min. Epitope-tagged proteins were
570 immunoprecipitated with either anti-Flag- or anti-HA-coated magnetic beads (Sigma and
571 Pierce, respectively, M8823 and 88836) following the manufacturer's instructions. RNA
572 was eluted from the beads with Proteinase K (Sigma, AM2546) and TRIzol LS Reagent
573 (Ambion, 10296028) was used to isolate immunoprecipitated RNA. The Ovation SoLo
574 RNA-Seq System (Tecan Genomics, S02240) was used to make PABP library. eCLIP
575 libraries were generated by following the protocol in Nostrand et al., 2016 (Van
576 Nostrand et al., 2016), with a minor modification that 8U of RNAse I (Ambion, AM2295)
577 was used per sample. Purified PABP CLIP libraries were submitted for paired-end 75bp
578 sequencing on the Illumina NextSeq platform at the Center for Genomics and
579 Bioinformatics (CGB) in Indiana University, Bloomington. eCLIP libraries were submitted
580 for paired-end 75bp sequencing on the Illumina NextSeq platform at the CGB in Indiana
581 University, Bloomington or paired-end 100bp sequencing on the DNBseq platform at the
582 BGI Genomics in China.
583
584 ***RNA-seq and PABP CLIP-seq data analysis***
585 RNA-seq datasets were obtained from the following references and are referred
586 to in the text by the surname of the first author: (Doupe et al., 2018; Dutta et al., 2015;
587 Fast et al., 2020; Korzelius et al., 2014; Leader et al., 2018; Li et al., 2018). Both RNA-

588 seq and PABP CLIP-seq read files were processed and aligned to the Berkeley
589 *Drosophila* Genome Project (BDGP) assembly release 6.28 (Ensembl release 99)
590 reference genome using a python based in house pipeline
591 (https://github.com/jkkbuddika/RNA-Seq-Data-Analyzer v1.0). Briefly, the quality of raw
592 sequencing files was assessed using FastQC (Andrews, 2010) version 0.11.9, low
593 quality reads were removed using Cutadapt (Martin, 2011) version 2.9, reads aligning to
594 rRNAs were removed using TagDust2 (Lassmann, 2015) version 2.2, remaining reads
595 were mapped to the Berkeley *Drosophila* Genome Project (BDGP) assembly release
596 6.28 (Ensembl release 99) reference genome using STAR (Dobin et al., 2013) genome
597 aligner version 2.7.3a and deduplicated using SAMtools (Li et al., 2009) version 1.10.
598 Subsequently, the aligned reads were counted to the nearest overlapping feature suing
599 the Subread (Liao et al., 2019) version 2.0.0 function *featureCounts*. Finally, bigWig files
600 representing RNA-seq coverage were generated using deepTools (Ramirez et al.,
601 2016) version 3.4.2 with the settings --normalizeUsing CPM --binSize 1. All programs,
602 versions and dependencies required to execute the RNA-seq data analyzer are
603 described in the user guide and can be installed using miniconda. The TMM-normalized
604 expression values were computed using the Bioconductor package edgeR (Robinson et
605 al., 2010) version 3.28.1. Genes with TMM-normalized expression value greater than 10
606 in all replicates were considered as a moderately expressed gene during transcriptome
607 comparisons. Differential gene expression analysis was performed with the
608 Bioconductor package DESeq2 (Love et al., 2014) version 1.26.0 using FDR < 0.05 and
609 $\log_2$ fold change > 1, unless otherwise noted. For both edgeR-based transcriptome
610 analysis and DESeq2-based differential gene expression analysis only protein coding

genes were used. All data visualization steps were performed using custom scripts written in R.

Enhanced-CLIP (eCLIP) data analysis

The eCLIP read file processing, alignment and processing was done using an easy to use python based in house analysis pipeline, eCLIP data analyzer (https://github.com/jkkbuddika/eCLIP-Data-Analyzer). While the eCLIP data analyzer pipeline permits analysis in both paired and single-end modes, we performed all our analyzes using the single-end mode, which uses only R2 read for analysis. The eCLIP data analyzer pipeline performs the following steps: (1) read quality assessment using FastQC (Andrews, 2010), (2) trimming of universal eCLIP adaptors using Cutadapt (Martin, 2011), (3) addition of UMI sequence (5'-NNNNNNNNNN in the R2 read) using UMI-tools (Smith et al., 2017) to read name to facilitate PCR duplicate removal, (4) removal of reads aligning to rRNAs using Tagdust2 (Lassmann, 2015), (5) alignment of remaining reads to the provided genome (i.e., BDGP assembly release 6.28, Ensembl release 100) using STAR (Dobin et al., 2013), (6) coordinating sort and index alignment outputs using SAMtools (Li et al., 2009), (7) removal of PCR duplicates using UMI-tools (Smith et al., 2017), (8) generation of bigwig files representing RNA-seq coverage tracks using deepTools (Ramirez et al., 2016), (9) quantification of nearest overlapping features using Subread (Liao et al., 2019) function *featureCounts*, and (10) peak calling using PureCLIP (Krakau et al., 2017). All programs, versions and dependencies required to execute the eCLIP data analyzer are described in the user guide and can be installed using miniconda. Input normalized peak calling output files were processed as

634 described in (Busch et al., 2020). Subsequently, significantly enriched *de novo* binding
635 motifs were identified using DREME (Bailey, 2011). Protein coding FMRP target genes
636 were used to identify enriched Gene Ontology (GO) terms using gProfiler. A selected
637 significantly enriched GO categories were plotted using R. Unless otherwise noted, all
638 data visualization steps were performed using custom scripts written in R.
639
640 ***CLIP-qPCR***
641 Intestine were dissected from 100 w^{1118} females, and FMRP was precipitated
642 with anti-FMRP antibody (5A11, DSHB) using the same method described above. RNA
643 was isolated from either the starting intestinal lysate or from the IP with Trizol LS
644 (Ambion, 10296028). Resulting RNA was treated with Turbo DNase(ThermoFisher ,
645 AM2239) and ~200ng of input RNA and all the RNA derived from IP sample was used
646 for cDNA synthesis with Superscript III (ThermoFisher, 56575). qPCR was performed
647 using the PowerUp SYBR Green Master Mix (ThermoFisher, A25742) on a
648 StepOnePlus machine (ThermoFisher). Primers for all targets detected are listed in
649 Table S5. Transcript levels were quantified in duplicate and normalized to *CG10116*.
650 Fold enrichment was calculated as the ratio of transcript in IP versus input.
651
652 ***Statistical analysis:***
653 Statistical analyses were performed using Prism (GraphPad, Version 7.0). Datasets
654 were tested for normality using D'Agostino-Pearson test. Comparisons of two datasets
655 was performed with an unpaired t-test while qPCR samples were analyzed with a paired
656 t-test. The statistical significance of overlaps in Venn Diagrams was calculated using a

657 hypergeometric test on R. When multiple pairwise comparisons were needed, the R
658 package GeneOverlap was used to perform Fischer's exact test which yields the
659 statistical significance, odds ratio and Jaccard indices for each pairwise comparison.
660 Significance is indicated as follows: n.s., not significant; *$p < 0.05$; **$p < 0.01$; ***$p <$
661 0.001; ****$p < 0.0001$.
662
663 ***Data availability:***
664 The PABP CLIP-seq and FMRP eCLIP-seq datasets from this study have been
665 submitted to the NCBI Gene Expression Omnibus under accession number XXXX (to
666 be added later).

Acknowledgements

We thank Steven Hou, Gerald Rubin, Steve Stowers, the Kyoto Drosophila Stock Center, the Bloomington Drosophila Stock Center (supported by grant NIHP4OOD018537), the Drosophila Genome Resource Center (supported by grant NIH2P40OD010949), and the Developmental Studies Hybridoma Bank for reagents; the Light Microscopy Imaging Center (supported by grant NIH1S10OD024988-01) for access to the SP8 confocal; and the National Institute of General Medical Sciences (Award R01GM124220) for financial support. The authors declare no competing financial interests.

Competing Interests

No competing interest declared.

680 **References:**

681

682 Andra, I., Ulrich, H., Durr, S., Soll, D., Henkel, L., Angerpointner, C., Ritter, J., Przibilla,
683 S., Stadler, H., Effenberger, M.*, et al.* (2020). An Evaluation of T-Cell Functionality After
684 Flow Cytometry Sorting Revealed p38 MAPK Activation. Cytometry A *97*, 171-183.

685

686 Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.
687 http://wwwbioinformaticsbabrahamacuk/projects/fastqc.

688

689 Andriatsilavo, M., Stefanutti, M., Siudeja, K., Perdigoto, C.N., Boumard, B., Gervais, L.,
690 Gillet-Markowska, A., Al Zouabi, L., Schweisguth, F., and Bardin, A.J. (2018). Spen
691 limits intestinal stem cell self-renewal. PLoS genetics *14*, e1007773.

692

693 Antonello, Z.A., Reiff, T., Ballesta-Illan, E., and Dominguez, M. (2015). Robust intestinal
694 homeostasis relies on cellular plasticity in enteroblasts mediated by miR-8-Escargot
695 switch. The EMBO journal *34*, 2025-2041.

696

697 Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data.
698 Bioinformatics *27*, 1653-1659.

699

700 Bardin, A.J., Perdigoto, C.N., Southall, T.D., Brand, A.H., and Schweisguth, F. (2010).
701 Transcriptional control of stem cell maintenance in the Drosophila intestine.
702 Development *137*, 705-714.

703

704 Barolo, S., Carver, L.A., and Posakony, J.W. (2000). GFP and beta-galactosidase
705 transformation vectors for promoter/enhancer analysis in Drosophila. Biotechniques *29*,
706 726, 728, 730, 732.

707

708 Baser, A., Skabkin, M., Kleber, S., Dang, Y., Gulculer Balta, G.S., Kalamakis, G.,
709 Gopferich, M., Ibanez, D.C., Schefzik, R., Lopez, A.S.*, et al.* (2019). Onset of
710 differentiation is post-transcriptionally controlled in adult neural stem cells. Nature *566*,
711 100-104.

712

713 Biteau, B., Karpac, J., Supoyo, S., Degennaro, M., Lehmann, R., and Jasper, H. (2010).
714 Lifespan extension by preserving proliferative homeostasis in Drosophila. PLoS
715 genetics *6*, e1001159.

716

717 Buddika, K., Ariyapala, I.S., Hazuga, M.A., Riffert, D., and Sokol, N.S. (2020a).
718 Canonical nucleators are dispensable for stress granule assembly in Drosophila
719 intestinal progenitors. Journal of cell science *133*.

720

721 Buddika, K., Huan, Y., Butrum-Griffith, A., Norrell, S., O'Connor, A., Patel, V., Rector,
722 S., Slovan, M., Sokolowsky, M., Kato, Y.*, et al.* (2020b). Intestinal progenitor P-bodies
723 maintain stem cell identity by suppressing pro-differentiation factors.
724 BioRxiv/2020/175398.

725

726 Busch, A., Bruggemann, M., Ebersberger, S., and Zarnack, K. (2020). iCLIP data
727 analysis: A complete pipeline from sequencing reads to RBP binding sites. Methods
728 *178*, 49-62.
729
730 Chen, C.H., Luhur, A., and Sokol, N. (2015). Lin-28 promotes symmetric stem cell
731 division and drives adaptive growth in the adult Drosophila intestine. Development *142*,
732 3478-3487.
733
734 Chen, X., and Dickman, D. (2017). Development of a tissue-specific ribosome profiling
735 approach in Drosophila enables genome-wide evaluation of translational adaptations.
736 PLoS genetics *13*, e1007117.
737
738 Cvetkovska, V., Hibbert, A.D., Emran, F., and Chen, B.E. (2013). Overexpression of
739 Down syndrome cell adhesion molecule impairs precise synaptic targeting. Nature
740 neuroscience *16*, 677-682.
741
742 Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone,
743 E.F., Chen, C., Fak, J.J., Chi, S.W.*, et al.* (2011). FMRP stalls ribosomal translocation
744 on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.
745
746 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
747 Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner.
748 Bioinformatics *29*, 15-21.

749

750 Doupe, D.P., Marshall, O.J., Dayton, H., Brand, A.H., and Perrimon, N. (2018).
751 Drosophila intestinal stem and progenitor cells are major sources and regulators of
752 homeostatic niche signals. Proceedings of the National Academy of Sciences of the
753 United States of America *115*, 12218-12223.
754

755 Dutta, D., Dobson, A.J., Houtz, P.L., Glasser, C., Revah, J., Korzelius, J., Patel, P.H.,
756 Edgar, B.A., and Buchon, N. (2015). Regional Cell-Specific Transcriptome Mapping
757 Reveals Regulatory Complexity in the Adult Drosophila Midgut. Cell reports *12*, 346-
758 358.
759

760 Fast, D., Petkau, K., Ferguson, M., Shin, M., Galenza, A., Kostiuk, B., Pukatzki, S., and
761 Foley, E. (2020). Vibrio cholerae-Symbiont Interactions Inhibit Intestinal Repair in
762 Drosophila. Cell reports *30*, 1088-1100 e1085.
763

764 Feng, H., Bao, S., Rahman, M.A., Weyn-Vanhentenryck, S.M., Khan, A., Wong, J.,
765 Shah, A., Flynn, E.D., Krainer, A.R., and Zhang, C. (2019). Modeling RNA-Binding
766 Protein Specificity In Vivo by Precisely Registering Protein-RNA Crosslink Sites.
767 Molecular cell *74*, 1189-1204 e1186.
768

769 Furriols, M., and Bray, S. (2001). A model Notch response element detects Suppressor
770 of Hairless-dependent molecular switch. Current biology : CB *11*, 60-64.
771

772 Greenblatt, E.J., and Spradling, A.C. (2018). Fragile X mental retardation 1 gene
773 enhances the translation of large autism-related proteins. Science *361*, 709-712.
774
775 Guo, Z., Driver, I., and Ohlstein, B. (2013). Injury-induced BMP signaling negatively
776 regulates Drosophila midgut homeostasis. The Journal of cell biology *201*, 945-961.
777
778 Hida, N., Aboukilila, M.Y., Burow, D.A., Paul, R., Greenberg, M.M., Fazio, M., Beasley,
779 S., Spitale, R.C., and Cleary, M.D. (2017). EC-tagging allows cell type-specific RNA
780 analysis. Nucleic acids research *45*, e138.
781
782 Hung, R.J., Hu, Y., Kirchner, R., Liu, Y., Xu, C., Comjean, A., Tattikota, S.G., Li, F.,
783 Song, W., Ho Sui, S.*, et al.* (2020). A cell atlas of the adult Drosophila midgut.
784 Proceedings of the National Academy of Sciences of the United States of America *117*,
785 1514-1523.
786
787 Hwang, H.W., Park, C.Y., Goodarzi, H., Fak, J.J., Mele, A., Moore, M.J., Saito, Y., and
788 Darnell, R.B. (2016). PAPERCLIP Identifies MicroRNA Targets and a Role of
789 CstF64/64tau in Promoting Non-canonical poly(A) Site Usage. Cell reports *15*, 423-435.
790
791 Jiang, H., and Edgar, B.A. (2009). EGFR signaling regulates the proliferation of
792 Drosophila adult midgut progenitors. Development *136*, 483-493.
793

794 Kim, J.H., Wang, X., Coolon, R., and Ye, B. (2013). Dscam expression levels determine
795 presynaptic arbor sizes in Drosophila sensory neurons. Neuron *78*, 827-838.
796
797 Korzelius, J., Azami, S., Ronnen-Oron, T., Koch, P., Baldauf, M., Meier, E., Rodriguez-
798 Fernandez, I.A., Groth, M., Sousa-Victor, P., and Jasper, H. (2019). The WT1-like
799 transcription factor Klumpfuss maintains lineage commitment of enterocyte progenitors
800 in the Drosophila intestine. Nat Commun *10*, 4123.
801
802 Korzelius, J., Naumann, S.K., Loza-Coll, M.A., Chan, J.S., Dutta, D., Oberheim, J.,
803 Glasser, C., Southall, T.D., Brand, A.H., Jones, D.L.*, et al.* (2014). Escargot maintains
804 stemness and suppresses differentiation in Drosophila intestinal stem cells. The EMBO
805 journal *33*, 2967-2982.
806
807 Krakau, S., Richard, H., and Marsico, A. (2017). PureCLIP: capturing target-specific
808 protein-RNA interaction footprints from single-nucleotide CLIP-seq data. Genome Biol
809 *18*, 240.
810
811 Lassmann, T. (2015). TagDust2: a generic method to extract reads from sequencing
812 data. BMC Bioinformatics *16*, 24.
813
814 Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A., and Dow, J.A.T. (2018). FlyAtlas 2:
815 a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-
816 Seq and sex-specific data. Nucleic acids research *46*, D809-D815.

817

818 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
819 Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The
820 Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
821

822 Li, H., and Jasper, H. (2016). Gastrointestinal stem cells in health and disease: from
823 flies to humans. Disease models & mechanisms *9*, 487-499.
824

825 Li, M., Shin, J., Risgaard, R.D., Parries, M.J., Wang, J., Chasman, D., Liu, S., Roy, S.,
826 Bhattacharyya, A., and Zhao, X. (2020). Identification of FMR1-regulated molecular
827 networks in human neurodevelopment. Genome research *30*, 361-374.
828

829 Li, Q., Nirala, N.K., Nie, Y., Chen, H.J., Ostroff, G., Mao, J., Wang, Q., Xu, L., and Ip,
830 Y.T. (2018). Ingestion of Food Particles Regulates the Mechanosensing Misshapen-
831 Yorkie Pathway in Drosophila Intestinal Growth. Developmental cell *45*, 433-449 e436.
832

833 Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster,
834 cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic
835 acids research *47*, e47.
836

837 Lin, G., Xu, N., and Xi, R. (2008). Paracrine Wingless signalling controls self-renewal of
838 Drosophila intestinal stem cells. Nature *455*, 1119-1123.
839

840 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
841 dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.
842
843 Luhur, A., Buddika, K., Ariyapala, I.S., Chen, S., and Sokol, N.S. (2017). Opposing
844 Post-transcriptional Control of InR by FMRP and LIN-28 Adjusts Stem Cell-Based
845 Tissue Growth. Cell reports *21*, 2671-2677.
846
847 Luhur, A., Chawla, G., and Sokol, N.S. (2013). MicroRNAs as components of systemic
848 signaling pathways in Drosophila melanogaster. Current topics in developmental biology
849 *105*, 97-123.
850
851 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput
852 sequencing reads. 2011 *17*, 3.
853
854 Maurin, T., Lebrigand, K., Castagnola, S., Paquet, A., Jarjat, M., Popa, A., Grossi, M.,
855 Rage, F., and Bardoni, B. (2018). HITS-CLIP in various brain areas reveals new targets
856 and new modalities of RNA binding by fragile X mental retardation protein. Nucleic acids
857 research *46*, 6344-6355.
858
859 McClelland, L., Jasper, H., and Biteau, B. (2017). Tis11 mediated mRNA decay
860 promotes the reacquisition of Drosophila intestinal stem cell quiescence. Developmental
861 biology *426*, 8-16.
862

863 McGuire, S.E., Mao, Z., and Davis, R.L. (2004). Spatiotemporal gene expression
864 targeting with the TARGET and gene-switch systems in Drosophila. Science's STKE :
865 signal transduction knowledge environment *2004*, pl6.
866
867 McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash,
868 M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of
869 RNA-Binding Proteins. Cell *165*, 742-753.
870
871 Micchelli, C.A., and Perrimon, N. (2006). Evidence that stem cells reside in the adult
872 Drosophila midgut epithelium. Nature *439*, 475-479.
873
874 Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after
875 cell lysis: implications for the interpretation of immunoprecipitation analyses. Rna *10*,
876 1692-1694.
877
878 Miller, E.D., Kahsai, L., Buddika, K., Dixon, M.J., Kim, B.Y., Calvi, B.R., Sokol, N.S.,
879 Hawley, R.S., and Cook, K.R. (2020). Identification and characterization of breakpoints
880 and mutations on Drosophila melanogaster balancer chromosomes. G3 (Submitted).
881
882 Monzo, K., Dowd, S.R., Minden, J.S., and Sisson, J.C. (2010). Proteomic analysis
883 reveals CCT is a target of Fragile X mental retardation protein regulation in Drosophila.
884 Developmental biology *340*, 408-418.
885

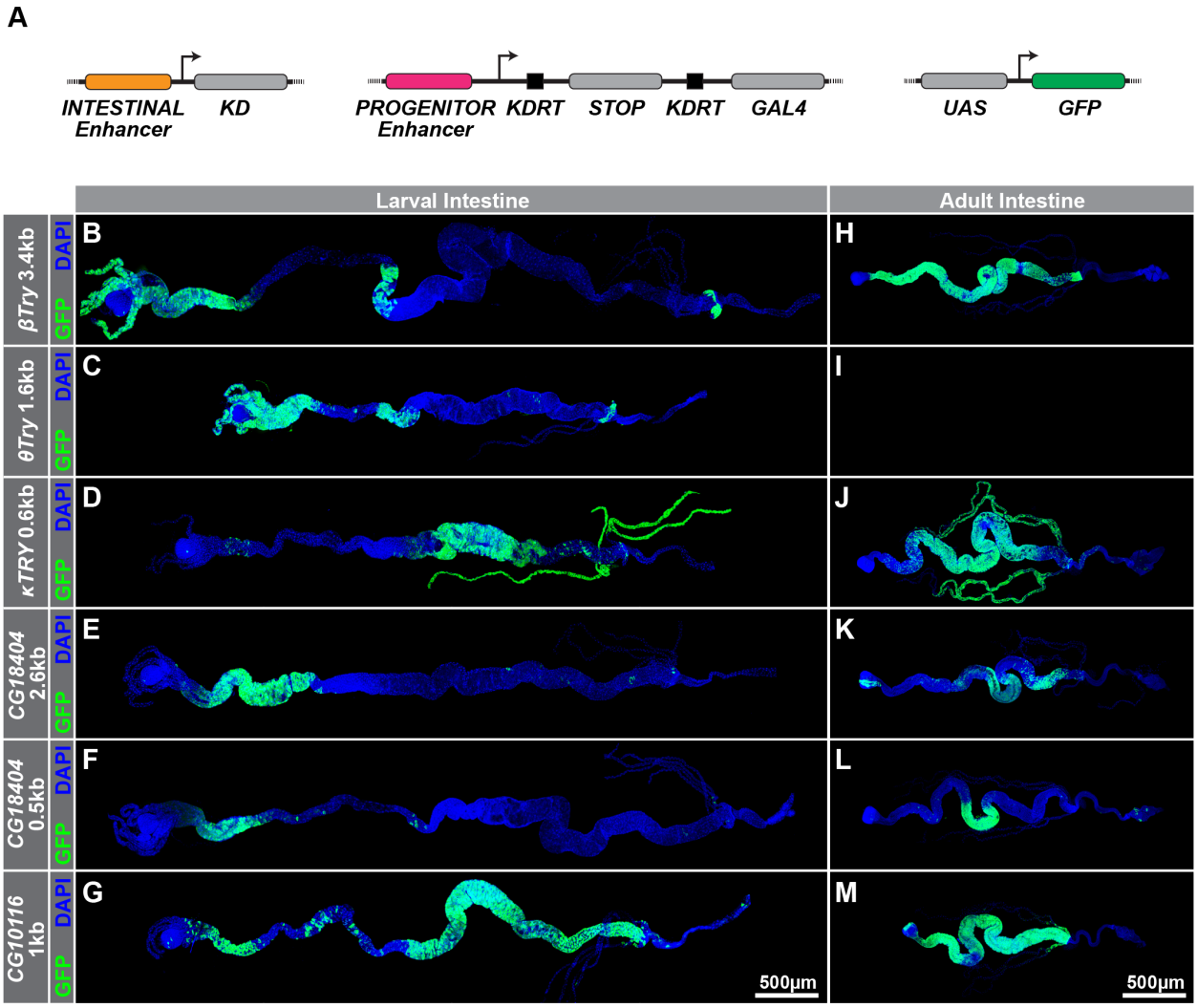
886 Nern, A., Pfeiffer, B.D., Svoboda, K., and Rubin, G.M. (2011). Multiple new site-specific
887 recombinases for use in manipulating animal genomes. Proceedings of the National
888 Academy of Sciences of the United States of America *108*, 14198-14203.
889
890 Phillips, M.D., and Thomas, G.H. (2006). Brush border spectrin is required for early
891 endosome recycling in Drosophila. Journal of cell science *119*, 1361-1370.
892
893 Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S.,
894 Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-
895 sequencing data analysis. Nucleic acids research *44*, W160-165.
896
897 Reeve, S.P., Bassetto, L., Genova, G.K., Kleyner, Y., Leyssen, M., Jackson, F.R., and
898 Hassan, B.A. (2005). The Drosophila fragile X mental retardation protein controls actin
899 dynamics by directly regulating profilin in the brain. Current biology : CB *15*, 1156-1163.
900
901 Richardson, G.M., Lannigan, J., and Macara, I.G. (2015). Does FACS perturb gene
902 expression? Cytometry A *87*, 166-175.
903
904 Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor
905 package for differential expression analysis of digital gene expression data.
906 Bioinformatics *26*, 139-140.
907

908 Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. (2002). Chromosomal clustering of
909 muscle-expressed genes in Caenorhabditis elegans. Nature *418*, 975-979.
910
911 Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., and Amieux, P.S. (2009). Cell-
912 type-specific isolation of ribosome-associated mRNA from complex tissues.
913 Proceedings of the National Academy of Sciences of the United States of America *106*,
914 13939-13944.
915
916 Sears, J.C., Choi, W.J., and Broadie, K. (2019). Fragile X Mental Retardation Protein
917 positively regulates PKA anchor Rugose and PKA activity to control actin assembly in
918 learning/memory circuitry. Neurobiol Dis *127*, 53-64.
919
920 Shearin, H.K., Macdonald, I.S., Spector, L.P., and Stowers, R.S. (2014). Hexameric
921 GFP and mCherry reporters for the Drosophila GAL4, Q, and LexA transcription
922 systems. Genetics *196*, 951-960.
923
924 Singh, S.R., Liu, W., and Hou, S.X. (2007). The adult Drosophila malpighian tubules are
925 maintained by multipotent stem cells. Cell stem cell *1*, 191-203.
926
927 Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in
928 Unique Molecular Identifiers to improve quantification accuracy. Genome research *27*,
929 491-499.
930

931 Sudhakaran, I.P., Hillebrand, J., Dervan, A., Das, S., Holohan, E.E., Hulsmeier, J.,
932 Sarov, M., Parker, R., VijayRaghavan, K., and Ramaswami, M. (2014). FMRP and
933 Ataxin-2 function together in long-term olfactory habituation and neuronal translational
934 control. Proceedings of the National Academy of Sciences of the United States of
935 America *111*, E99-E108.
936
937 Tahmasebi, S., Amiri, M., and Sonenberg, N. (2018). Translational Control in Stem
938 Cells. Front Genet *9*, 709.
939
940 Tenenbaum, S.A., Carson, C.C., Lager, P.J., and Keene, J.D. (2000). Identifying mRNA
941 subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proceedings
942 of the National Academy of Sciences of the United States of America *97*, 14085-14090.
943
944 Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y.,
945 Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K.*, et al.* (2016). Robust
946 transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP
947 (eCLIP). Nature methods *13*, 508-514.
948
949 Voght, S.P., Fluegel, M.L., Andrews, L.A., and Pallanck, L.J. (2007). Drosophila NPC1b
950 promotes an early step in sterol absorption from the midgut epithelium. Cell metabolism
951 *5*, 195-205.
952

953 Wang, X., Mu, Y., Sun, M., and Han, J. (2017). Bidirectional regulation of fragile X
954 mental retardation protein phosphorylation controls rhodopsin homoeostasis. J Mol Cell
955 Biol *9*, 104-116.
956
957 Yang, Z., Edenberg, H.J., and Davis, R.L. (2005). Isolation of mRNA from specific
958 tissues of Drosophila by mRNA tagging. Nucleic acids research *33*, e148.
959
960 Zeng, X., Chauhan, C., and Hou, S.X. (2010). Characterization of midgut stem cell- and
961 enteroblast-specific Gal4 lines in drosophila. Genesis *48*, 607-611.
962
963 Zhang, Y.Q., Bailey, A.M., Matthies, H.J., Renden, R.B., Smith, M.A., Speese, S.D.,
964 Rubin, G.M., and Broadie, K. (2001). Drosophila fragile X-related gene regulates the
965 MAP1B homolog Futsch to control synaptic structure and function. Cell *107*, 591-603.
966
967

968 **Figures and Legends:**



969

970 **Figure 1: *KD* driver analysis identifies a pan-intestinal enhancer in adults**. (A)
971 Schematic of the I-KCKT system, which involves three transgenes: (i) a KD-expressing
972 transgene under the control of an intestinal enhancer, (ii) a transgene in which a KDRT-
973 flanked stop-cassette separates GAL4 from a progenitor enhancer, and (iii) a UAS
974 responder, in this case, controlling GFP expression. (B-M) KD recombinase-mediated
975 labeling of (B-G) larval and (H-M) adult intestinal cells. KD lines contain (B,H) 3.4kb
976 *βTry*, (C, I) 1.6kb *θTry*, (D, J) 0.6kb *κTry*, (E, K) 2.6kb *CG18404*, (F, L) 0.5kb *CG18404*,
977 or (G, M) 1kb *CG10116* enhancer sequences. KD expression pattern was detected

978 based on KDRT-mediated activation of a *tubulin*-based *GAL4.p65* driver in combination

979 with a *UAS-6XGFP* responder. Full genotypes are listed in Table S4.

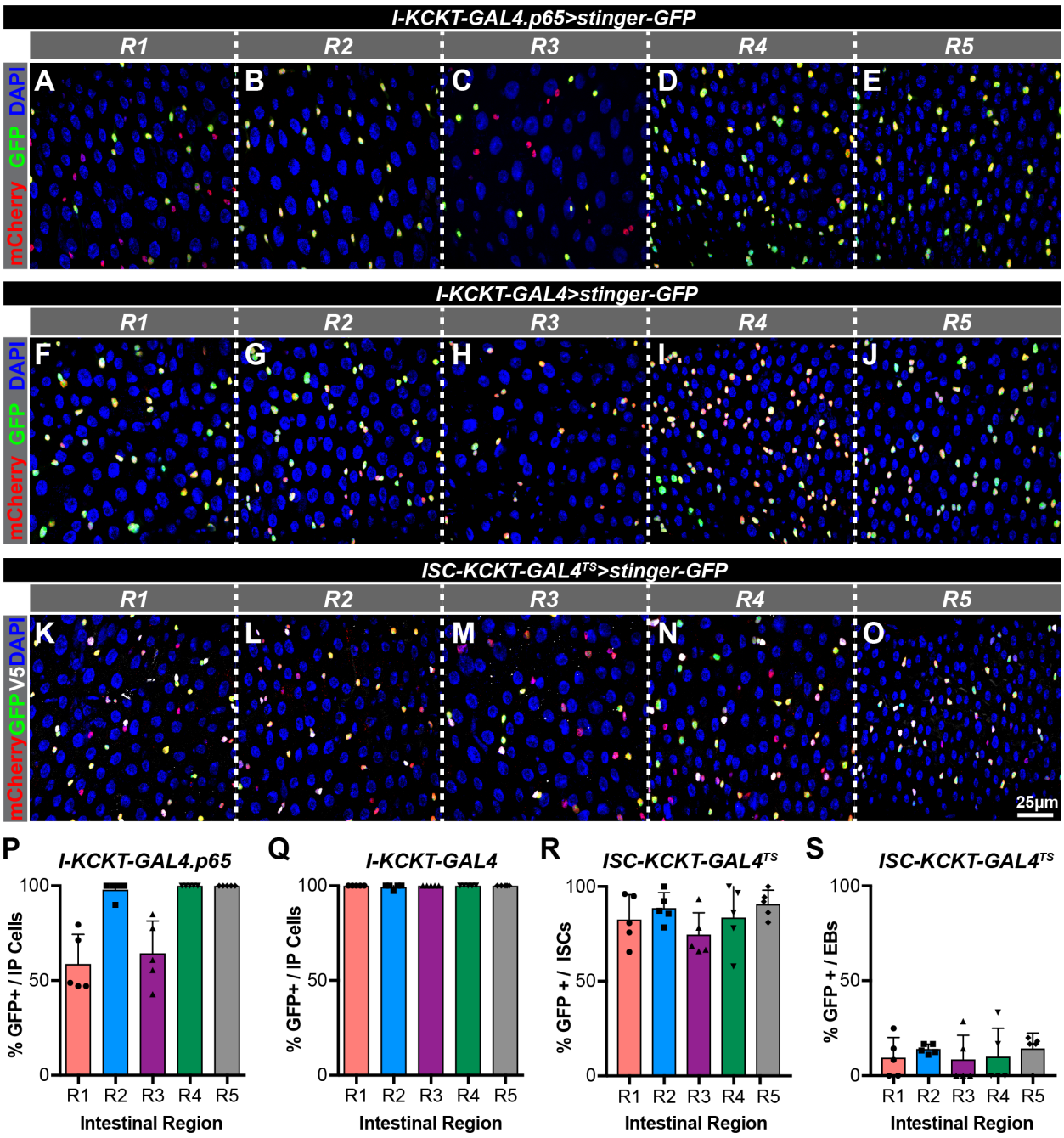
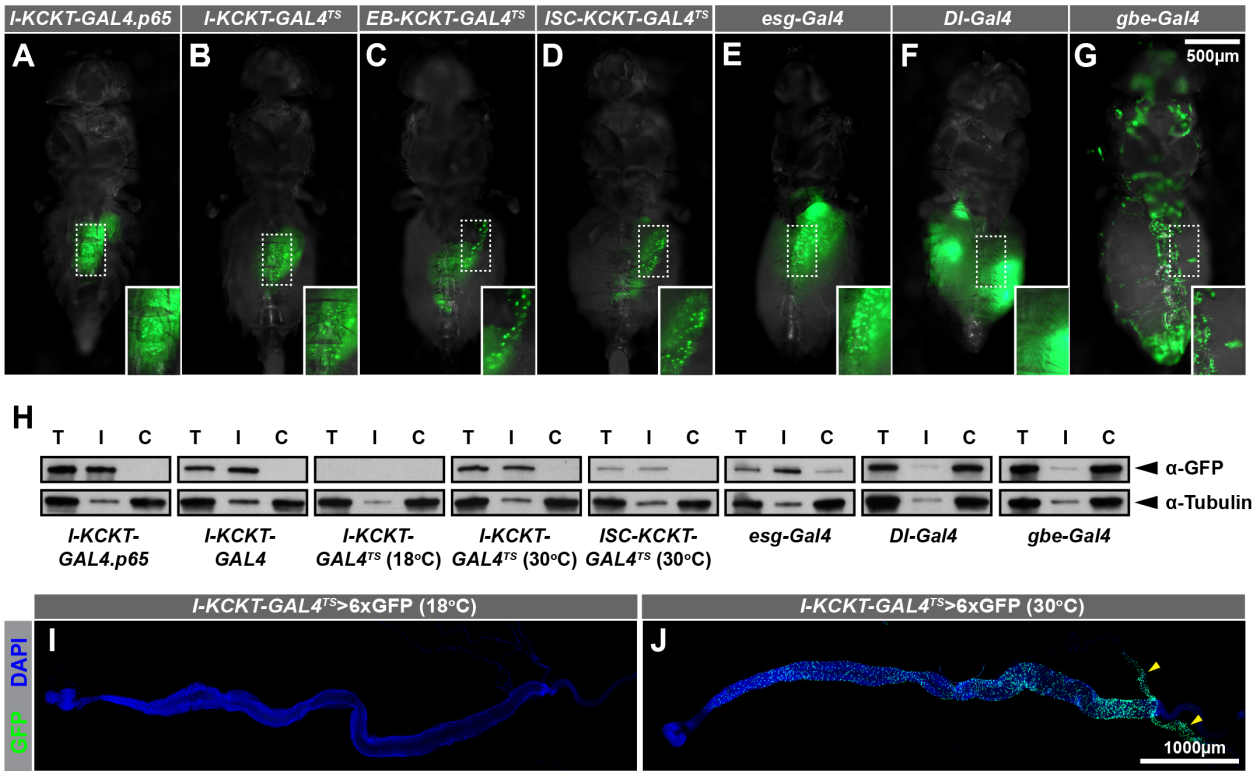


Figure 2: *I-KCKT-GAL4.p65* and *I-KCKT-GAL4* label most intestinal progenitor cells and *ISC-KCKT-GAL4^{TS}* labels most ISCs. (A-O) Micrographs of five intestinal regions (R1-R5) stained for *stinger-GFP* in green, the intestinal progenitor marker *mira-His2A.mCherry.HA* in red, and the DAPI DNA marker in blue. GFP expression is driven by either (A-E) *I-KCKT-GAL4.p65*, (F-J) *I-KCKT-GAL4*, or (K-O) *ISC-KCKT-GAL4^{TS}*. (P-

986 Q) Histograms showing percentage of *mira-His2A.mCherry.HA*-labeled intestinal
987 progenitor (IP) cells per intestinal region (R1-R5) that are labeled with *stinger-GFP*
988 driven by either (A-E) *I-KCKT-GAL4.p65* or (F-J) *I-KCKT-GAL4*. (R-S) Histograms
989 showing the percentage of (R) ISCs and (S) EBs labeled with *stinger-GFP* driven by (K-
990 O) *ISC-KCKT-GAL4*^{TS}. Full genotypes are listed in Table S4.

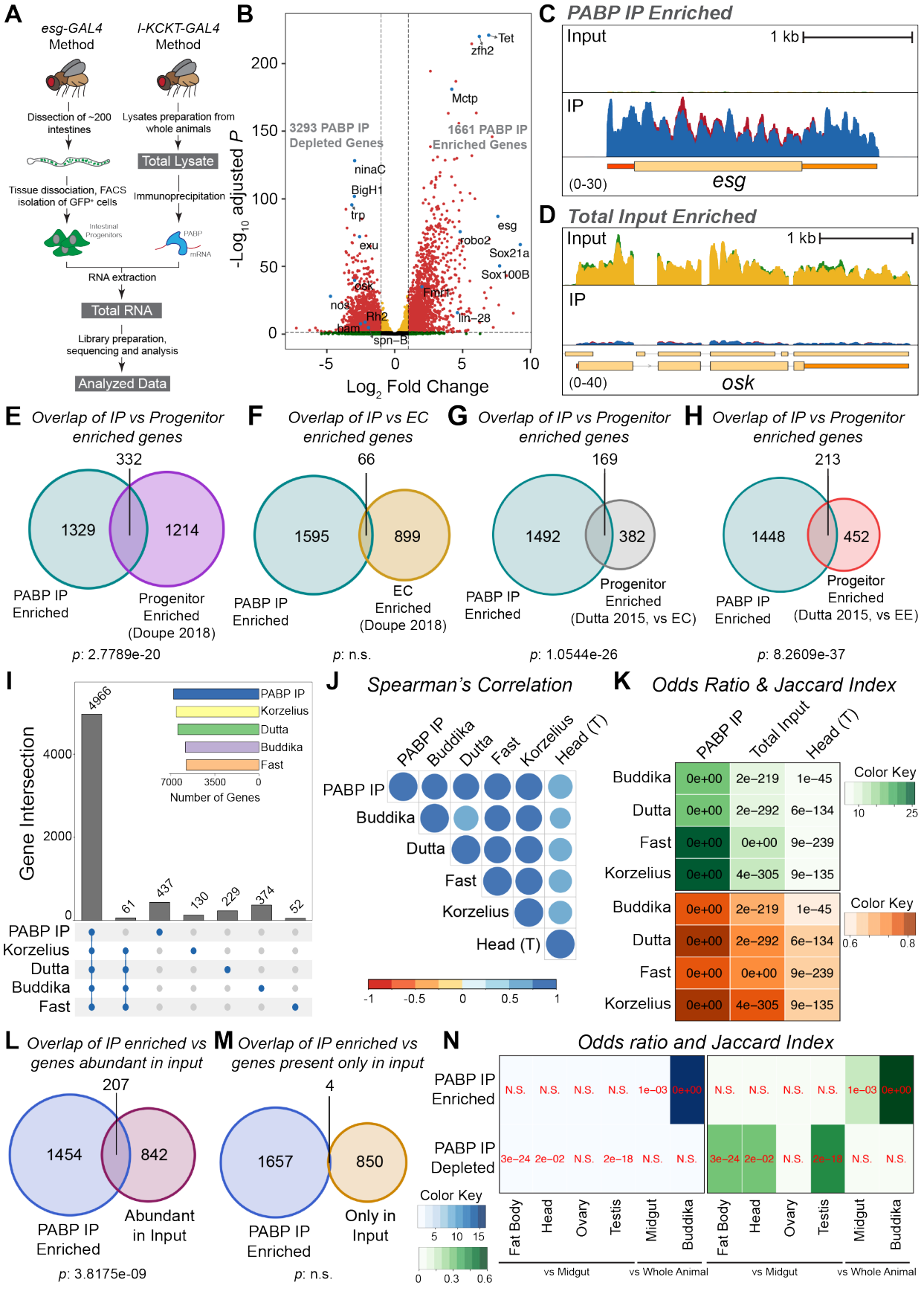


991

992 **Figure 3: *I-KCKT-GAL4* drivers are intestine-specific**. (A-G) Pictures of adult females
993 showing *6XGFP* expression patterns driven by (A) *I-KCKT-GAL4.p65*, (B) *I-KCKT-*
994 *GAL4*^{TS}, (C) *EB-KCKT-GAL4*^{TS}, (D) *ISC-KCKT-GAL4*^{TS}, (E) *esg*$^{P\{GawB\}NP5130}$*-GAL4*, (F)
995 *Dl*$^{05151\text{-}G}$*-GAL4*, or (G) *Su(H)GBE-GAL4*. Insets are enlargements of abdominal regions
996 displaying intestinal GFP expression. (H) Western blot analysis of tissue extracts from
997 total animals (T), dissected intestines (I), or dissected carcasses (C) probed with either
998 anti-GFP (top) or anti-Tubulin (bottom) antibodies. Lysates were generated from adults
999 harboring *UAS-6XGFP* driven by *I-KCKT-GAL4.p65*, I-*KCKT-GAL4*, I-*KCKT-GAL4*^{TS} at
1000 18°C, I-*KCKT-GAL4*^{TS} at 30°C, *ISC-KCKT-GAL4*^{TS} at 30°C, *esg*$^{P\{GawB\}NP5130}$*-GAL4*,
1001 *Dl*$^{05151\text{-}G}$*-GAL4*, or *Su(H)GBE-GAL4*. (I, J) Micrographs of intestines showing *6XGFP*
1002 expression patterns driven by *I-KCKT-GAL4*^{TS} from adults kept either at (I) 18°C or (J)
1003 30°C and counterstained for the DAPI DNA marker in blue. Yellow arrowheads indicate

1004 GFP expression at the base of the malphigian tubules. Full genotypes are listed in

1005 Table S4.


A
esg-GAL4 Method
I-KCKT-GAL4 Method
Dissection of ~200 intestines
Lysates preparation from whole animals
Total Lysate
Tissue dissociation, FACS isolation of GFP+ cells
Immunoprecipitation
Intestinal Progenitors
PABP
mRNA
RNA extraction
Total RNA
Library preparation, sequencing and analysis
Analyzed Data
B
3293 PABP IP Depleted Genes
1661 PABP IP Enriched Genes
-Log10 adjusted P
Log2 Fold Change
Tet
zfh2
Mctp
ninaC
BigH1
trp
exu
esg
robo2
Sox21a
Sox100B
osk
nos
Rh2
bam
spn-B
Fmr1
lin-28
C PABP IP Enriched
Input
IP
1 kb
(0-30)
esg
D Total Input Enriched
Input
IP
1 kb
(0-40)
osk
E Overlap of IP vs Progenitor enriched genes
332
1329
1214
PABP IP Enriched
Progenitor Enriched (Doupe 2018)
p: 2.7789e-20
F Overlap of IP vs EC enriched genes
66
1595
899
PABP IP Enriched
EC Enriched (Doupe 2018)
p: n.s.
G Overlap of IP vs Progenitor enriched genes
169
1492
382
PABP IP Enriched
Progenitor Enriched (Dutta 2015, vs EC)
p: 1.0544e-26
H Overlap of IP vs Progenitor enriched genes
213
1448
452
PABP IP Enriched
Progeitor Enriched (Dutta 2015, vs EE)
p: 8.2609e-37
I
Gene Intersection
Number of Genes
PABP IP
Korzelius
Dutta
Buddika
Fast
4966
61
437
130
229
374
52
J Spearman's Correlation
PABP IP
Buddika
Dutta
Fast
Korzelius
Head (T)
K Odds Ratio & Jaccard Index
PABP IP
Total Input
Head (T)
Buddika 0e+00 2e−219 1e−45
Dutta 0e+00 2e−292 6e−134
Fast 0e+00 0e+00 9e−239
Korzelius 0e+00 4e−305 9e−135
Color Key
L Overlap of IP enriched vs genes abundant in input
207
1454
842
PABP IP Enriched
Abundant in Input
p: 3.8175e-09
M Overlap of IP enriched vs genes present only in input
4
1657
850
PABP IP Enriched
Only in Input
p: n.s.
N Odds ratio and Jaccard Index
PABP IP Enriched
PABP IP Depleted
Fat Body
Head
Ovary
Testis
Midgut
Buddika
vs Midgut
vs Whole Animal
Color Key

1006

1007 **Figure 4: CLIP-seq of PABP driven by *I-KCKT-GAL4* identifies progenitor**
1008 **expressed genes.** (A) Schematic representation of conventional FACS (left) and *I-*
1009 *KCKT*-based (right) intestinal progenitor RNA-sequencing. (B) Volcano plot of
1010 differentially expressed genes in PABP IP versus total input. Each dot represents a
1011 single gene. Yellow indicates a false discovery rate adjusted p-value (FDR) < 0.05 and
1012 a Log_2 fold change < 1 or > -1. Green indicates Log_2 fold change > 1 or < -1 and an
1013 FDR >= 0.05. Red indicates an FDR < 0.05 and log_2 fold change > 1 or < -1. A selected
1014 set of significantly changed genes have been labeled in blue. (C-D) Genome browser
1015 tracks of normalized total input and PABP IP at the (C) *esg* or (D) *osk* loci. Note that the
1016 two replicates of the total input (green and yellow) and the two replicates of the PABP IP
1017 (red and blue) have been overlayed. (E-H) Venn diagrams showing genes significantly
1018 enriched in PABP-CLIP that overlap with Doupe's progenitor enriched (E) genes, (F)
1019 Doupe's EC enriched genes, (G) Dutta's progenitor enriched genes as compared to EC
1020 genes, or (H) Dutta's progenitor enriched genes as compared to EE genes. (I) Upset
1021 plot showing the overlap of moderately expressed genes identified by PABP CLIP-seq
1022 versus RNA-seq of FACS-isolated progenitor cells reported in Korzelius, Dutta, Buddika
1023 and Fast. Numbers above each bar shows the size of each intersection. (J)
1024 Correlogram of Spearman's *rho* values for pairwise comparisons of progenitor
1025 expressed genes in PABP CLIP-seq, progenitor RNA-seq reported in Buddika, Dutta,
1026 Fast, or Korzelius, or RNA-seq from *Drosophila* female head. (K) Heatmaps of odds
1027 ratio (in green) and Jaccard index (in orange) values for pairwise comparisons of PABP
1028 IP CLIP-seq, RNA-seq of total input, or RNA-seq of dissected female heads against
1029 FACS-based progenitor RNA-seq. Numbers on each colored box show the p-value for

1030 each overlap based on Fisher's exact test. (L-M) Venn diagrams showing genes
1031 significantly enriched in PABP-CLIP that overlap with (L) the top 10% of input genes
1032 based on normalized expression values, or (M) genes present in input but not in any of
1033 the FACS-based RNA-seq gene lists. (N) Heatmaps of odds ratio (in blue) and Jaccard
1034 index (in green) values of pairwise comparisons of PABP IP enriched or PABP IP
1035 depleted genes to six other gene sets. These sets include fat body, head, ovary or testis
1036 enriched genes (identified relative to midgut genes) as well as midgut or progenitor
1037 enriched genes from Buddika (identified relative to whole animal input). Numbers on
1038 each colored box show the p-value for each overlap based on Fisher's exact test.
1039

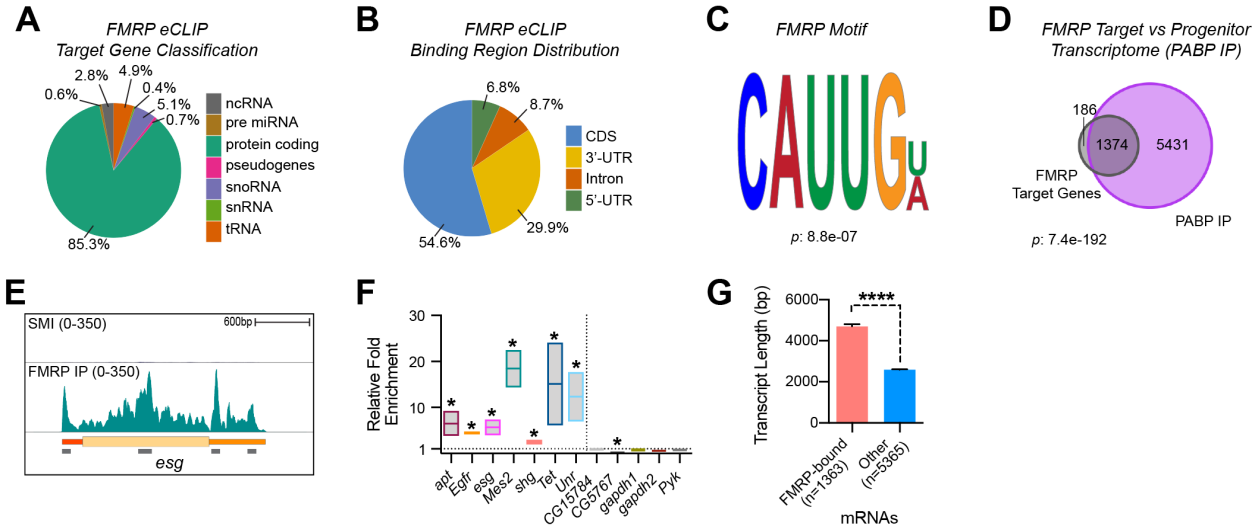
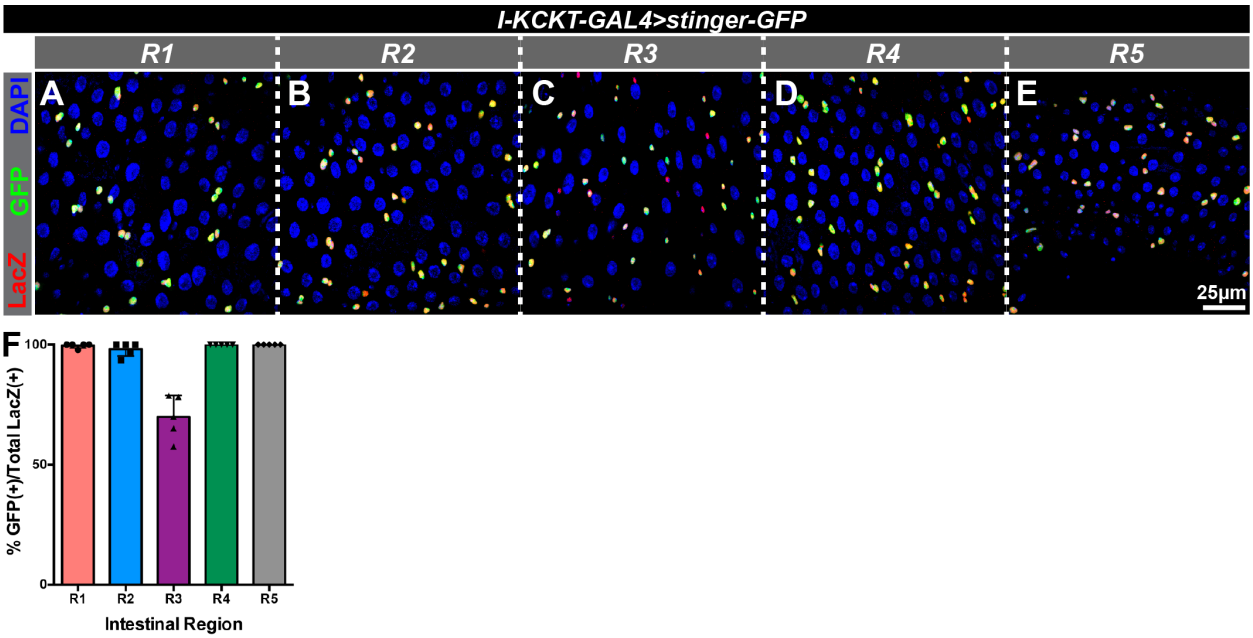


Figure 5: eCLIP of FMRP driven by *I-KCKT-GAL4* identifies intestinal target mRNAs. (A-B) Pie chart representing the percentage of (A) target gene types and (B) binding site distribution in mRNAs identified in FMRP eCLIP. (C) The top FMRP binding motif identified using DREME. (D) Venn diagram showing overlap between FMRP target genes with the progenitor transcriptome (identified by PABP IP). (E) Genome browser tracks of normalized SMI and FMRP IP at the *esg* locus of the *Drosophila* genome. Note that the two replicates of FMRP IP were merged prior to genome browser visualization. Locations of four FMRP binding regions are shown with grey bars. (F) Bar plot showing fold enrichment of twelve mRNAs in FMRP IP compared to whole intestinal input as determined by qPCR. Negative controls are separated from the other genes by a dotted line. (G) Bar plot showing the mean transcript length of FMRP-bound versus non-bound transcripts that are expressed in intestinal progenitors. Note that the longest transcript isoform was used for this analysis whenever a gene had multiple transcript isoforms. *p < 0.05; ****p < 0.0001.

1056 **Supplementary Figure Legends:**

1057



1058

1059 **Figure S1: *I-KCKT-GAL4* and *esg-LacZ* label the same cells**. (A-E) Micrographs of

1060 five intestinal regions (R1-R5) stained for *I-KCKT-GAL4*-driven *stinger-GFP* in green,

1061 the intestinal progenitor marker *esg-LacZ* in red, and the DAPI DNA marker in blue. (F)

1062 Histogram of the percentage of *esg-LacZ*-positive cells that also -*stinger-GFP* in each

1063 region.

1064

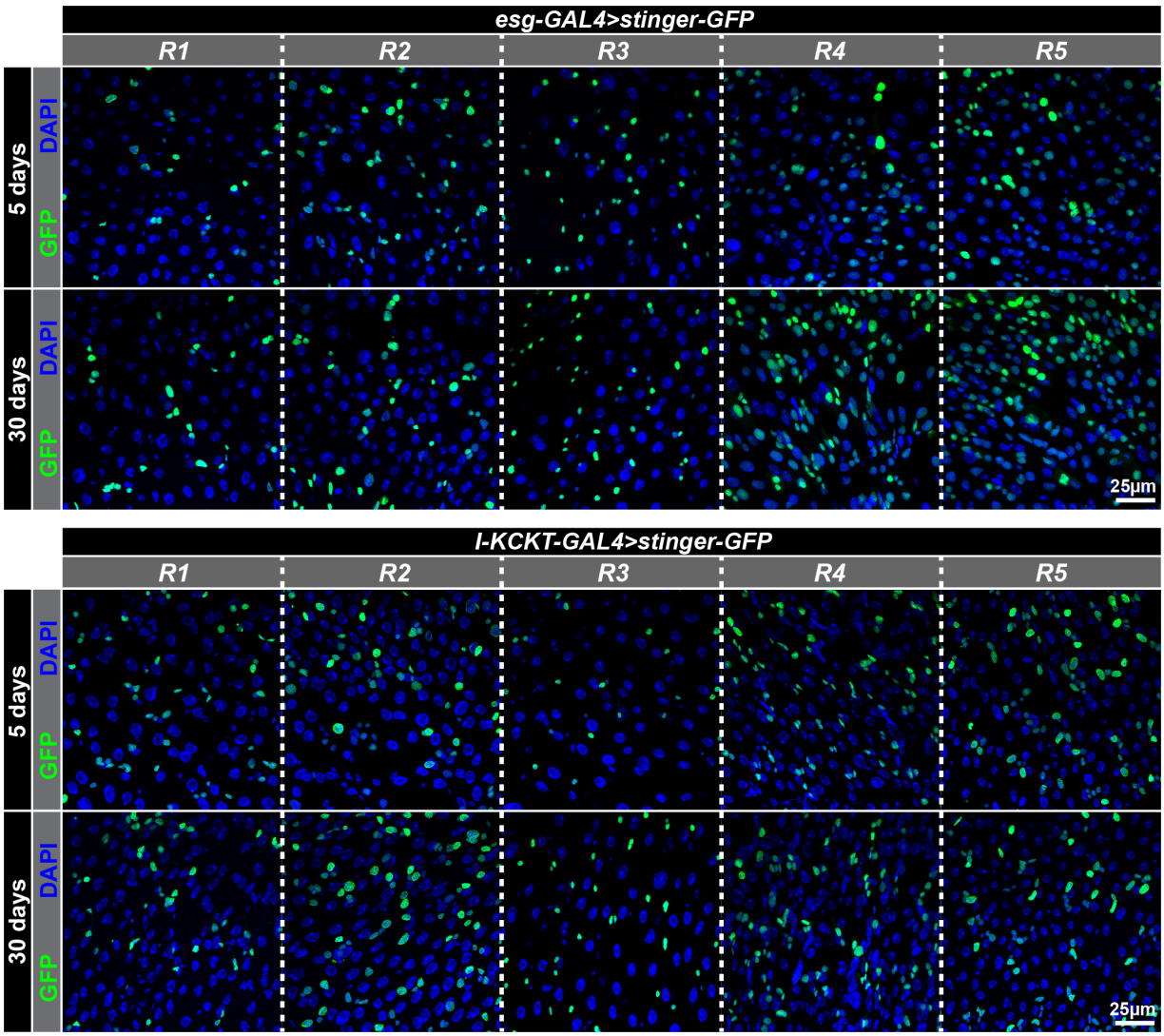
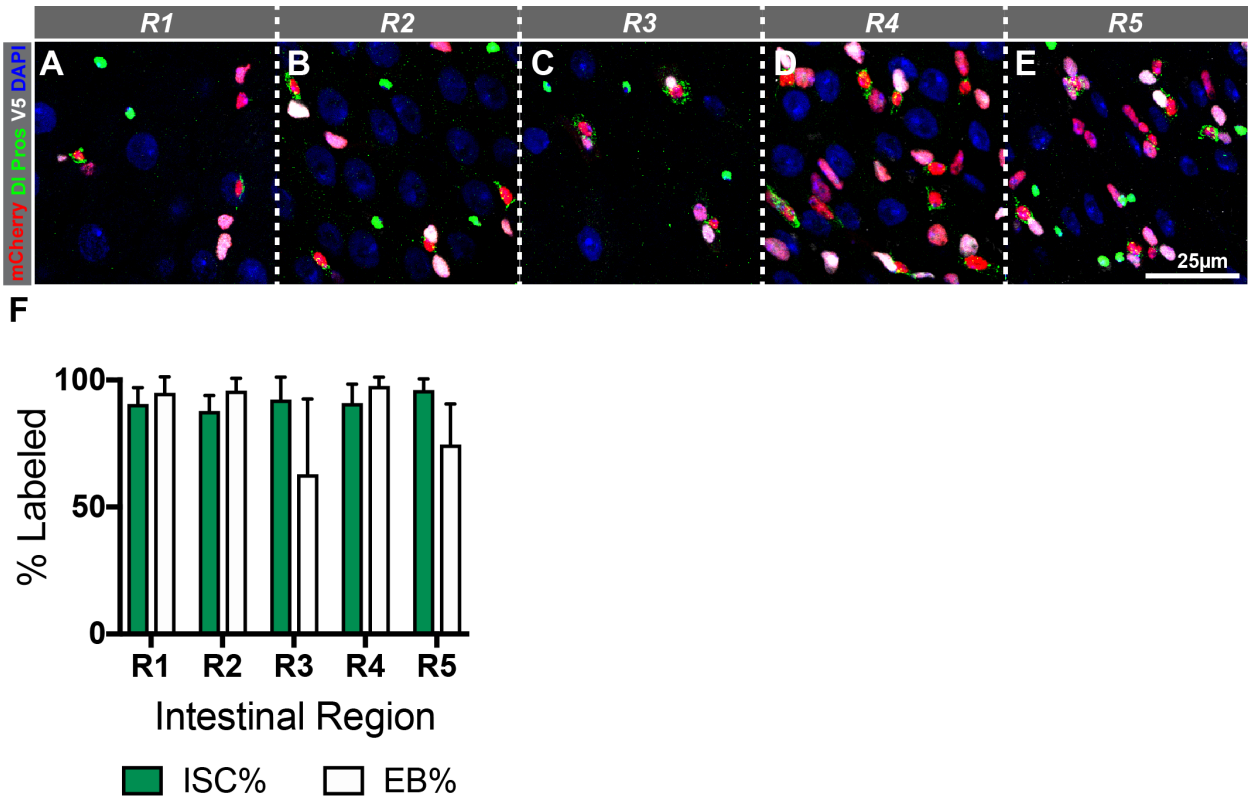


Figure S2: Comparison of *I-KCKT-GAL4* and *esg-GAL4* expression in aged flies. Representative micrographs of the five intestinal regions (R1-R5) of young (5 day) and older (30-day) females showing *UAS-stinger-GFP* driven by either *esg-GAL4* (top) or *I-KCKT-GAL4.p65* (bottom) and counterstained for the DAPI DNA marker in blue.



1071

1072 **Figure S3: Verification that dual reporters *mira-His2A.mCherry.HA* and *3Xgbe-***

1073 ***smGFP.V5.nls* distinguish ISCs and EBs.** (A-E) Representative micrographs of the

1074 five regions (R1-R5) from the intestines of 5-day old females stained for ISCs (anti-

1075 Delta, green cytoplasmic staining), EEs (anti-Prospero, green nuclear staining), *mira-*

1076 *His2A.mCherry.HA* (red), *3Xgbe-smGFP.V5.nls (white),* and the DAPI DNA marker

1077 (blue). (F) Quantification of the percent of ISCs and EBs labeled by the dual reporters

1078 *mira-His2A.mCherry.HA* and *3Xgbe-smGFP.V5.nls* in each of the five intestinal regions.

1079 ISCs were defined as the number of Delta+ cells per field of view. EBs were defined as

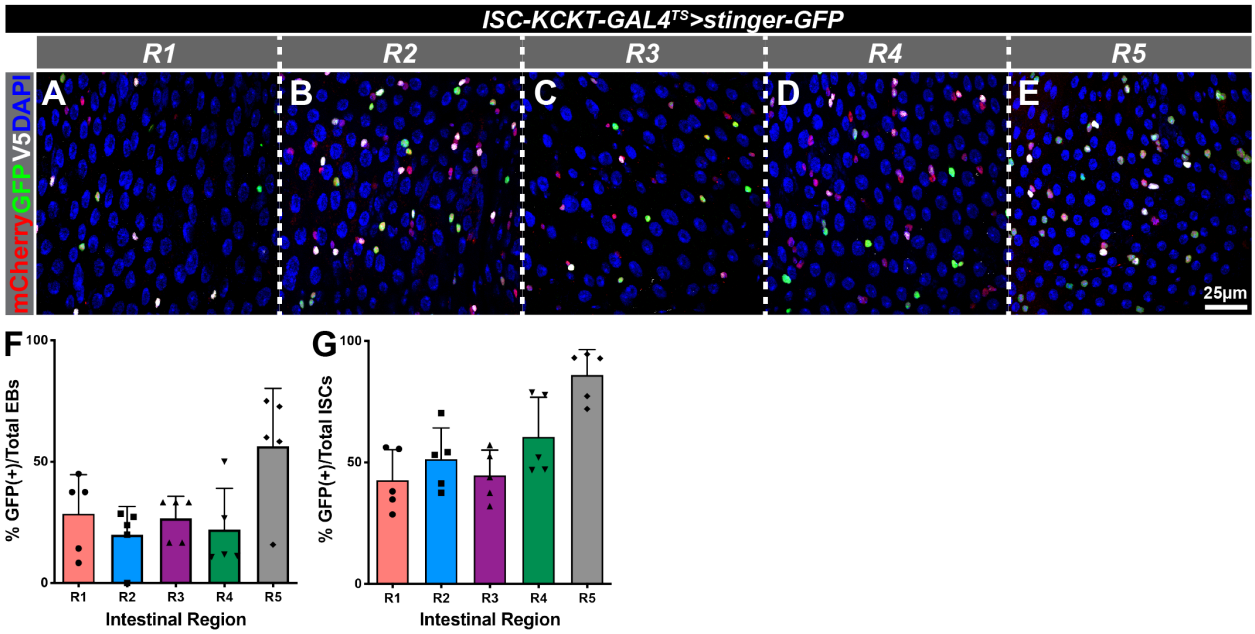
1080 the number of Delta-, mCherry+ cells per field of view. For quantification, field of view

1081 images were taken and quantified from defined portions of each intestinal region in 6-7

1082 intestines. We note that we detected a few (1-5) cells that were both Delta+ and *3Xgbe-*

1083 *smGFP.V5.nls*+ in regions of most intestines. These were not scored as either ISCs or

1084 EBs, and could represent cells transitioning between these two cell fates.



1085

1086 **Figure S4: *ISC-KCKT-GAL4^{TS}* labels many ISCs.** (A-E) Representative micrographs of

1087 the five regions (R1-R5) from the intestines of 5-day old females stained for *ISC-KCKT-*

1088 *GAL4^{TS}*-driven *UAS-stinger-GFP* in green, *mira-His2A.mCherry.HA* in red, *3Xgbe-*

1089 *smGFP.V5.nls* in white, and the DAPI DNA marker in blue. ISCs were scored as cells

1090 positive for *mira-His2A.mCherry.HA* and negative for *3Xgbe-smGFP.V5.nls*, while EBs

1091 were scored as cells positive for both markers. (F) Histogram of the percentage of EBs

1092 that expressed *ISC-KCKT-GAL4^{TS}*-driven *UAS-stinger-GFP* in each region. (G)

1093 Histogram of the percentage of ISCs that expressed *ISC-KCKT-GAL4^{TS}*-driven *UAS-*

1094 *stinger-GFP* in each region.

1095

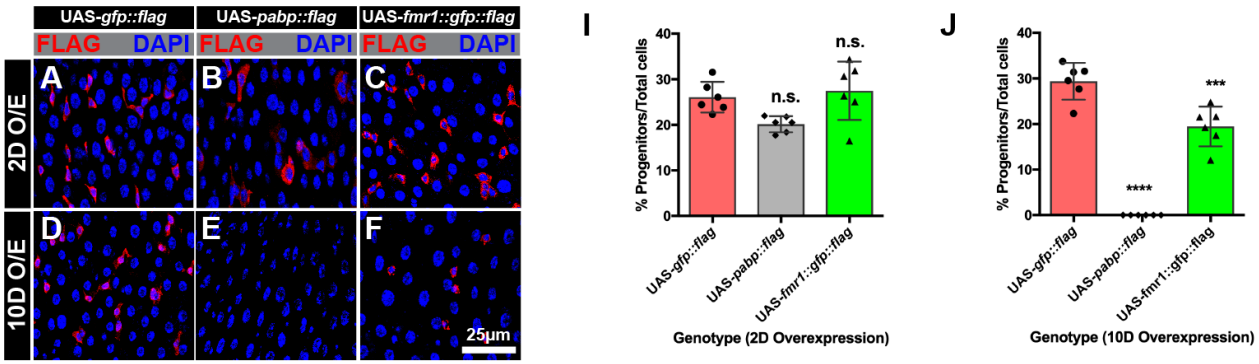


Figure S5: *I-KCKT-GAL4^{TS}*-dependent expression of tagged RBPs in intestinal progenitor cells. (A-J) Micrographs of intestines showing *I-KCKT-GAL4^{TS}*-dependent expression of *UAS-GFP.FLAG* (A, D), *UAS-PABP.FLAG* (B,E), and *UAS-FMR1.GFP.FLAG* (C,F) after two (A-C) and ten (D-F) days at the permissive temperature. Expressed proteins were detected with either anti-FLAG or -HA antibodies (shown in red) and intestines were counterstained with the DAPI DNA marker in blue. Full genotypes are listed in Table S4. (I, J) Histograms of the percentage of progenitor cells after two and ten days of *I-KCKT-GAL4^{TS}*-dependent expression, based on quantification of the numbers of HA+ or FLAG^{+} and DAPI^{+} cells.

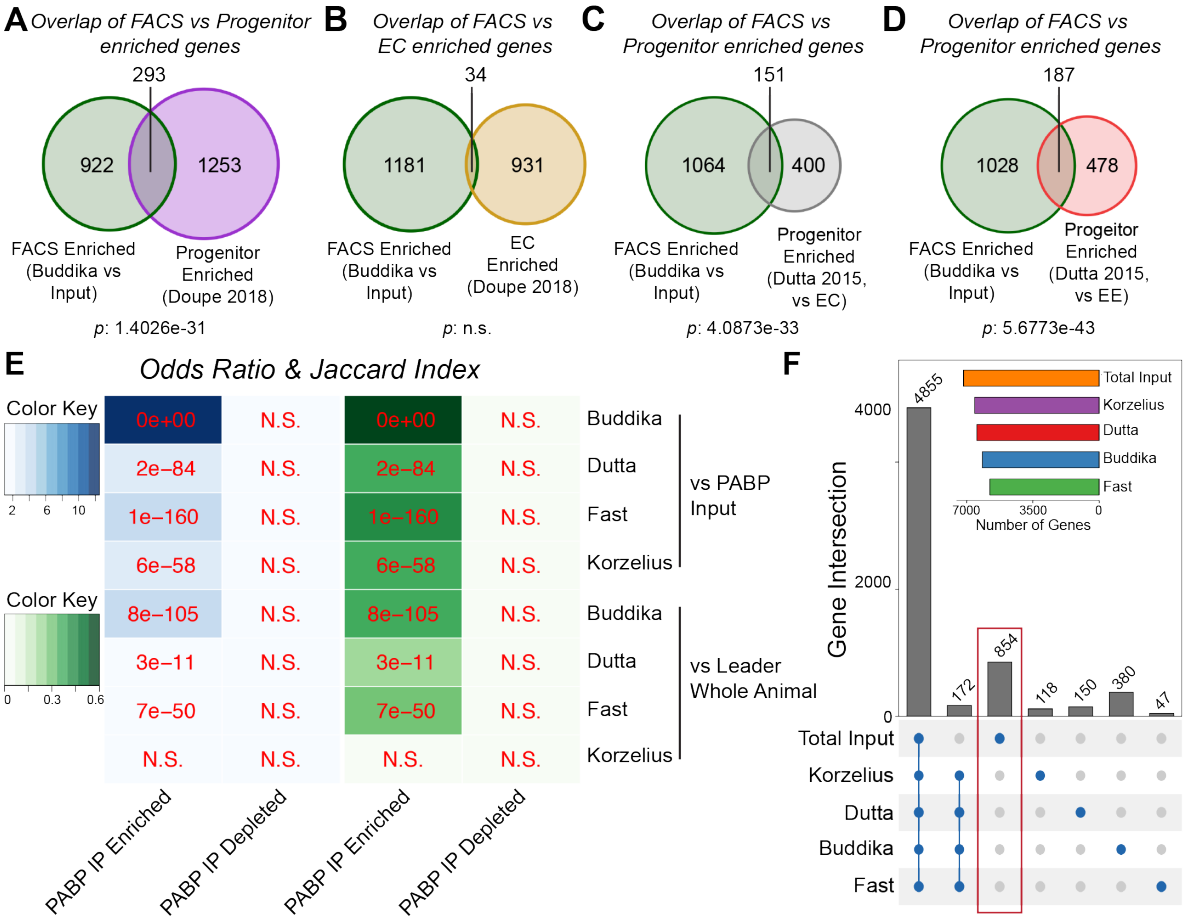


Figure S6: I-KCKT based RIPs identifies genes expressed in intestinal progenitors. (A-D) Venn diagrams showing genes significantly enriched in FACS-isolated intestinal progenitor data from Buddika that overlap with Doupe's progenitor enriched genes (A), Doupe's EC enriched genes (B), Dutta's progenitor enriched genes as compared to EC genes (C), or Dutta's progenitor enriched genes as compared to EE genes (D). (E) Heatmaps of odds ratio (in blue) and Jaccard index (in green) values for pairwise comparisons of PABP IP enriched or PABP IP depleted genes against input normalized FACS-based progenitor RNA-seq. Two sources of input data have been used for FACS data normalization: total input of the PABP experiment, or the female whole animal dataset from Leader. Numbers on each colored box show the p-value for

1117 each overlap based on Fisher's exact test. N.S.: Not significant. (F) Upset plot showing
1118 the overlap of moderately expressed genes identified by FACS based methods and total
1119 PABP IP input. Only a select set of meaningful overlaps are shown. The number above
1120 each bar shows the precise size of each intersection. Note that 854 genes were
1121 uniquely present in the input, but not in any of the FACS based datasets

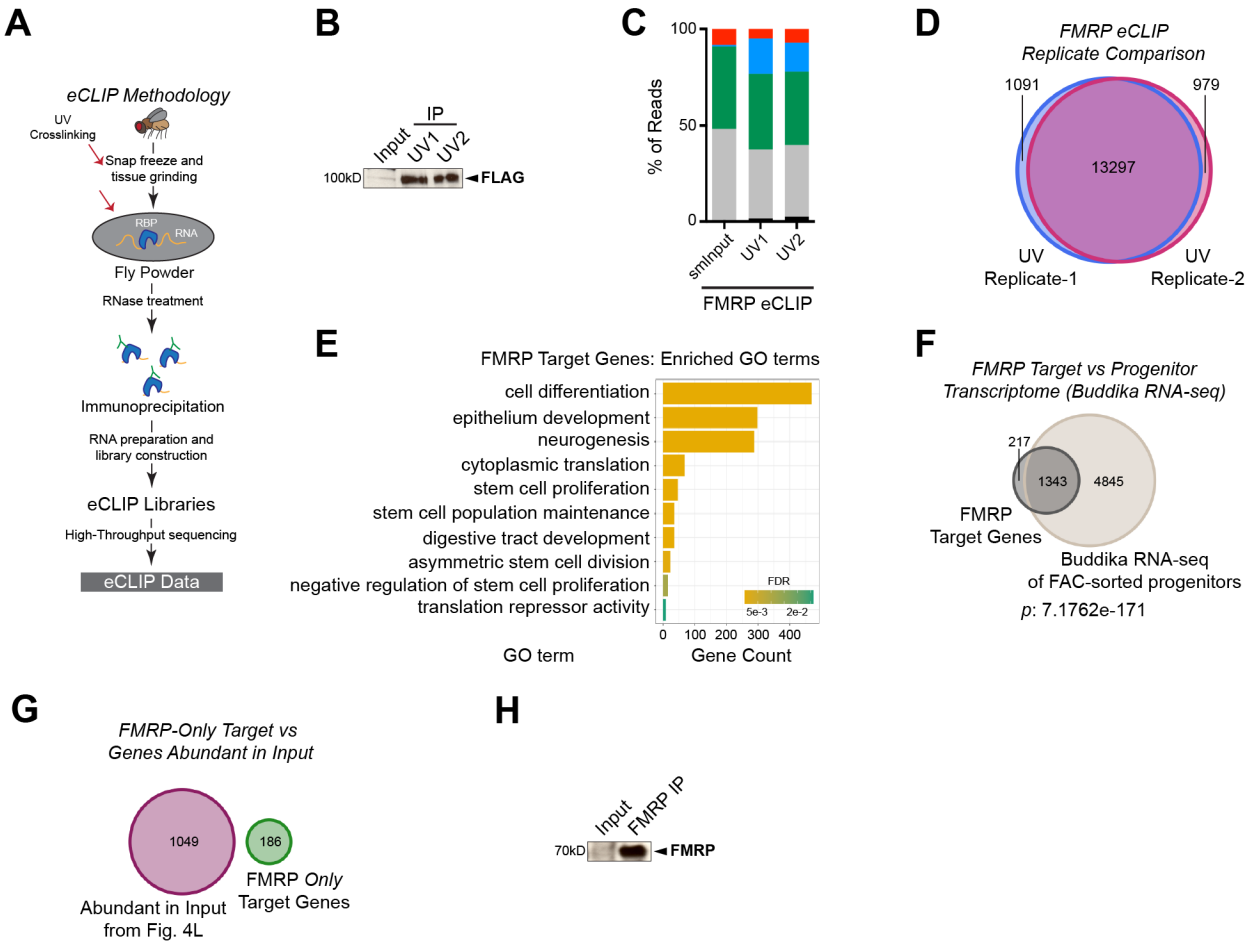


Figure S7: FMRP eCLIP using *I-KCKT* identifies target transcripts. (A) Schematic representation of the eCLIP methodology. (B) Western blot of input and FLAG IP samples (UV1 and UV2) prepared from whole animal extract and probed with anti-FLAG antibodies. (C) Grouped bar plot showing the percentage of sequencing read distribution. (D) Venn diagram showing the overlap of eCLIP peaks identified in UV replicate 1 and 2. (E) Bar plot showing a selected set of significantly enriched biological process Gene Ontology (GO) terms for FMRP target genes. (F) Venn diagram of FMRP target genes and intestinal progenitor transcriptome identified by RNA-seq of Buddika. (G) Venn diagram of top 10% genes with highest TMM normalized expression values from Fig. 4L and FMRP target genes that are absent in the progenitor transcriptome

1133 (186 genes from Fig. 5D). Note that the 186 genes are unlikely to be contaminants as
1134 they do not overlap with most highly abundant genes in the preparation. (H) Western
1135 blot of input and FMRP IP samples prepared from dissected intestines probed with anti-
1136 FMRP antibodies.
1137

1138 **Supplementary Tables**

1139

1140 Descriptions of the five supplementary Tables, supplied as excel files, are included

1141 below.

1142

1143 Table S1: List of genes differentially expressed in PABP IP versus whole animal Input.

1144 Table S2: List of genes common to either PABP IP or FACS-isolated progenitors

1145 (Buddika 2020b) and DamID-identified progenitor or EC specific genes (Doupe et al.,

1146 2018).

1147 Table S3: List of genes genes identified by FMRP eCLIP.

1148 Table S4. List of fly genotypes in figures.

1149 Table S5. List of DNA primers.