Transcription factors drive opposite relationships between gene age and tissue specificity in male and female *Drosophila* gonads

Evan Witt, Nicolas Svetec, Sigi Benjamin, Li Zhao*

Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY 10065, USA

*Correspondence to: lzhao@rockefeller.edu

Abstract

Evolutionarily young genes are usually preferentially expressed in the testis across species. While it is known that older genes are generally more broadly expressed than younger genes, the properties that shaped this pattern are unknown. Older genes may gain expression across other tissues uniformly, or faster in certain tissues than others. Using *Drosophila* gene expression data, we confirmed previous findings that younger genes are disproportionately testis-biased and older genes are disproportionately ovary-biased. We found that the relationship between gene age and expression is stronger in the ovary than any other tissue, and weakest in testis. We performed ATAC-seq on *Drosophila* testis and found that while genes of all ages are more likely to have open promoter chromatin in testis than in ovary, promoter chromatin alone does not explain the ovary-bias of older genes. Instead, we found that upstream transcription factor (TF) expression is highly predictive of gene expression in ovary, but not in testis. In ovary, TF expression is more predictive of gene expression than open promoter chromatin, whereas testis gene expression is similarly influenced by both TF expression and open promoter chromatin. We propose that the testis is uniquely able to expresses younger genes controlled by relatively few TFs, while older genes with more TF partners are broadly expressed with peak expression most likely in ovary. The testis allows widespread baseline expression that is relatively unresponsive to regulatory changes, whereas the ovary transcriptome is more responsive to trans-regulation and has a higher ceiling for gene expression.

Introduction

For eons, genes have continuously arisen by a multitude of ways, from duplication and divergence to *de novo* origination from non-genic DNA (Begun et al., 2006; Long et al., 2003; Ohno, 1970; Tautz and Domazet-Lošo, 2011; Zhao et al., 2014; Zhou et al., 2008). Gene birth and death is a continuous and dynamic process in evolution, culled by natural selection or genetic drift (Kaessmann, 2010; Palmieri et al., 2014). A large portion of young genes segregate within or recently fixed in populations, and most young genes are expressed specifically in the testis (Levine et al., 2006; Zhao et al., 2014), similar to duplicated genes (Long et al., 2013). The phrase "out of the testis" was originally used to describe young retroposed genes (Vinckenbosch et al., 2006), which gained expression by exploiting cis-regulatory machinery of nearby genes. Testis-bias has since been observed in young X-linked duplicate genes, leading researchers to propose that young genes escape Meiotic Sex Chromosome Inactivation (MSCI) due to immature cis-regulatory machinery (Zhang et al., 2010a). Testis expresses more genes in general than any other tissue (Soumillon et al., 2013), and studies from many taxa support that a large proportion of young genes then to show testis-biased or testis-specific expression and function (see review in Long et al., 2013).

The testis-biased expression of young genes has many possible explanations. Besides the obvious hypothesis that genes expressed in reproductive tissues may directly influence reproductive success and fitness (Begun et al., 2006; Zhang et al., 2004), many propose that the testis has a permissive chromatin environment facilitating the transcriptional birth of genetic novelties (Kaessmann, 2010; Soumillon et al., 2013). Indeed, most genes are at least somewhat expressed in the testis (Soumillon et al., 2013; Witt et al., 2019). It has long been proposed that an upregulation of universal transcriptional machinery facilitates such widespread transcription (Schmidt, 1996). Such broad transcription may be a form of genomic surveillance, meant to detect and repair mutations via transcription-coupled repair or other mechanisms (Grive et al., 2019; Xia et al., 2020). It has also been proposed that permissive testis transcription is also due to reduced mRNA degradation of testis-specific genes (Mayr, 2016). Young genes may also have low levels of "active" epigenetic markers across tissues, despite high expression in testis (Zhang and Zhou, 2019). Results from Zhang and Zhou 2019 suggest that young genes have similar epigenetic profiles across tissues, yet show testis-biased expression, while older genes show consistently higher levels of "active" epigenetic marks. Their results indicate that the "out of the

63 testis" pattern for the emergence of young genes may not be driven by specific epigenetic marks,
64 but rather by a context-dependent trans-regulatory environment between tissues (Ding et al.,
65 2010). Alternatively, recruitment of nearby testis-biased cis-regulatory elements by young genes
66 may also be responsible for many testis-biased new genes (Majic and Payne, 2020).
67 While it is known that young genes are often testis-specific, and that older genes are more
68 broadly-expressed than young genes (Kondo et al., 2017; Zhou et al., 2008), it is unknown how
69 this relationship works. When genes age, do older genes lose expression in testis, and retain
70 relatively constant expression in other tissues? Or do older genes maintain relatively constant
71 expression in testis, and gain expression in other tissues? If so, are all non-testis tissues equally
72 conducive to old genes, or do the genomic characteristics of older genes produce higher
73 expression in certain tissues? Once out of the testis, is any tissue the next hot target of tissue-
74 biased expression when the genes expand their functions in other tissues?
75 One clue is that older duplicated genes are more likely to be retained if they are ovary-
76 biased (Assis, 2019). This might imply a specific importance of older genes to ovary expression
77 and function. To this effect, researchers have identified several modules of highly conserved,
78 older genes with heightened importance in human ovarian function (Zhang et al., 2019). To see if
79 the *Drosophila* ovary drives the shift away from testis-bias in older genes, we analyzed a
80 database of RNA-seq data from FlyAtlas2 (Leader et al., 2018) to characterize tissue-bias for
81 genes of all ages in every tissue. We found that ovary has the largest relationship between gene
82 age and expression, explaining why the oldest genes are often ovary biased. Conversely, testis
83 shows a weaker relationship between gene age and gene expression than any other tissue.
84 To explain this trend, we examined the tissue-specific activity of the transcription factor
85 (TF) regulators of every gene in the DroID database (Murali et al., 2011). We found that ovary-
86 biased genes tend to have higher upstream TF expression than testis-biased genes of all age
87 groups, yet young genes, with fewer TF partners, tend to be testis-expressed and old genes, with
88 more TF partners, tend to be ovary biased. We found evidence that testis allows higher
89 transcription than the ovary for genes with low TF expression. Conversely, genes with high TF
90 expression have higher expression in the ovary than the testis. Additional upstream TF
91 expression appears to confer diminishing returns on expression in testis, but greatly benefits
92 ovary expression, explaining why older genes with more TF partners tend to be ovary biased.

93 After establishing the different relationships between trans regulation and gene
94 expression in testis and ovary, we performed ATAC-seq to assess if open promoter chromatin is
95 equally predictive of expression in the two tissues. All age groups of genes are more likely to
96 have open promoter chromatin in testis than ovary, indicating that open chromatin by itself is
97 insufficient to explain age-related expression bias. In ovary, we found that high upstream TF
98 expression is much more predictive of gene expression than the presence of open promoter
99 chromatin, whereas in testis, high TF expression and open promoter chromatin are similarly
100 predictive of gene expression. This indicates that gene expression in ovary is much more linked
101 to trans-regulatory factors than testis expression. Taken together with our observation that young
102 genes are less likely to be bound by annotated TFs than older genes, the opposite trends of gene
103 age and tissue bias in testis and ovary make biological sense. We published a web app to allow
104 users to interactively explore our tissue specificity data for any set of genes without coding
105 experience necessary: https://zhao.labapps.rockefeller.edu/tissue-specificity/.
106
107 **Results**
108 **Testis and ovary show an opposite relationship between gene age and tissue bias**
109 Using gene ages divided into Drosophilid (youngest), pre-Drosophilid (middle-aged), and
110 pre-Bilateria (oldest), and tissue RNA-seq data from FlyAtlas2, we find results consistent with
111 earlier work showing that younger genes are more tissue specific than older genes (Figure 1A).
112 We plotted the proportion of genes from each age group with maximum expression in testis,
113 ovary, and male and female carcasses with the reproductive tracts removed. A plurality of young
114 genes are testis biased, but the abundance of testis-biased genes declines for older genes (Figure
115 1B). Surprisingly, we found the opposite trend for ovary: older genes are very likely to have
116 maximum expression in ovary, but almost no younger genes are ovary biased. No other tissues
117 displayed a relationship of this magnitude (Supplemental Figure 1), indicating that the two
118 tissues that contribute most to gene age-related expression patterns are the male and female
119 reproductive tissues. Whereas young genes are often testis-biased and highly tissue specific, old
120 genes are broadly expressed with peak expression in ovary.
121 While a plurality of old genes are ovary-biased, this is not due to an increased likelihood
122 of expression for old genes in ovary. Young genes are most commonly expressed with FPKM>2
123 in testis (65%) and least commonly expressed in ovary (13%), whereas testis, ovary and somatic

124 tissues express a similar proportion of old genes (all between 73% and 85%; Figure 1C,
125 Supplemental Figure 2). Therefore, the age-related decline in testis-bias is not due to an absence
126 of old gene expression in testis. The proportion of genes expressed between age groups varies the
127 least in testis, and the most in ovary, indicating that ovary may have a disproportionately large
128 relationship between gene age and expression. We confirmed that young duplicate genes were
129 not confounding these results by repeating the analysis from Figure 1 with *melanogaster*-specific
130 genes removed (Supplemental Figure 3). We also confirmed these results with an alternate set of
131 gene age assignments (Supplemental Figure 4).
132

133 **Testis shows a weak, and ovary shows a strong relationship between gene age and**
134 **expression**

135 We wanted to further unpack how gene expression correlates with gene age across tissues
136 to understand our observed patterns of testis-bias and ovary-bias. For each tissue, we plotted
137 gene expression (Log2(FPKM +1)) from FlyAtlas2 conditioning by gene age. In every tissue,
138 expression of old genes was higher for pre-Bilateria genes than for Drosophilid genes as
139 measured with a pairwise Wilcoxon test (Figure 2A). In every tissue except testis, Drosophilid
140 genes were less expressed on average than pre-Drosophilid genes. In testis, these two groups
141 were statistically similar. This may be because in testis, unlike other tissues, a similar proportion
142 of genes are expressed for each age group (Figure 1C). A qualitative comparison shows that
143 expression of the 3 age groups is least different in testis (median FPKM 8.53 (Drosophilid), 3.19
144 (Pre-Drosophilid), 8.24 (pre-Bilateria)), and most dramatically different in ovary (median FPKM
145 0.071, 0.25, 21.10 respectively) (Figure 2A).

146 To quantitatively compare tissue-specific gene expression as a function of age group, we
147 performed a one-way ANOVA on each tissue and age group from Figure 2A. The ANOVA F
148 statistic is the ratio of between group variation to intra-group variation. For similar groups, the F
149 statistic is close to 1. The ANOVA F statistic is highest in ovary, meaning that the age groups are
150 more variable in this tissue than any other. In testis, the F statistic is lower, meaning that gene
151 expression varies less between age groups. Young genes have relatively similar expression in
152 testis across all age groups, in contrast to other tissues where gene expression is highly stratified
153 across age groups, with young genes the least and old genes the most expressed.

154 For each tissue, we also calculated the summed pairwise mean differences between every
155 group. This measure is the absolute value of the difference between the mean of each age group
156 within a tissue, summed for each pair of groups (Drosophilid vs. Pre-Drosophilid, Pre-
157 Drosophilid vs. Bilateria, Drosophilid vs. Bilateria). By this measure, mean testis expression is
158 the least different between gene age groups and ovary expression varies the most of any tissue
159 (Figure 2B). The results in Figure 2B hold if *melanogaster*-specific genes are removed
160 (Supplemental Figure 5), or with an alternate method of gene age assignments (Supplemental
161 Figure 6).
162
163 **Testis expression requires lower transcription factor activity than ovary expression**
164 We hypothesized that TFs may play a role in the discrepancy between age/expression
165 relationships between the testis and ovary. We designed a proxy measure of TF network activity
166 for every gene in every tissue. For every gene with bound TFs listed by DroID (Murali et al.,
167 2011), we defined the summed scaled expression of the upsteam TFs of a gene in a tissue as "TF
168 expression". Higher TF expression in a tissue indicates that a gene's TF partners are more
169 transcriptionally active in that tissue. This metric is based on data from ChIP-seq and ChIP-chip
170 experiments for individual TFs and only considers whether a TF binds to a given gene's
171 promoter. While such a method does not reveal whether a TF-gene relationship is one of
172 activation or repression, it is unbiased with regard to gene age since the whole-genome binding
173 profile of a TF is agnostic to the degree of study a particular gene has received (as young genes
174 are often less studied than older genes with mammalian homologs).
175 The purpose of our TF expression metric is not to infer gene expression (for which RNA-
176 seq is much better suited), but rather to assess the relative dynamics between gene expression
177 and trans regulation across tissues. For this purpose, the metric performs consistently well across
178 tissues even though some TFs are repressive in nature. For more details about TF expression, see
179 methods.
180 We compared the TF expression of young, middle-aged, and older genes between the
181 testis and ovary. We thought that since young genes are more specifically expressed in testis,
182 young genes would have higher uptsream TF expression in testis than in ovary. We found that no
183 age group of genes shows higher TF expression in testis than ovary (Figure 3A). The testis-
184 specificity of young genes must be due to factors other than increased TF expression in testis.

185 Exploring further, we found that young genes have fewer identified TF-gene interactions than
186 middle-aged genes, which in turn have fewer TF binding partners than old genes (Figure 3B).
187 We confirmed these results using an alternate list of gene ages in Supplemental Figure 7.
188 We then sought to correlate expression with TF activity between testis and ovary, and
189 found that genes with low TF expression are much more active in testis than in ovary.
190 Conversely, genes with high TF expression are often more active in the ovary than in testis
191 (Figure 3C). It appears that testis expression requires fewer TFs than ovary expression,
192 explaining why young genes, with fewer TFs, would have testis-biased expression. Having many
193 TF partners, a property of older genes, appears to boost expression in ovary more than in testis.
194 To confirm that this property was not sex-specific we compared TF expression and gene
195 expression in the male and female brain, two sexually dimorphic tissues, and observed no major
196 differences (Figure 3D). Additionally, we made this comparison across all tissues in FlyAtlas2
197 (Supplemental Figure 8), and found that gene expression is least correlated to TF expression in
198 testis (Pearson's r=0.22), and most responsive in ovary (Pearson's r=0.67).
199
200 **Testis promoter chromatin is broadly open across all gene ages**
201 To see whether promoter chromatin environment explains TF expression differences in
202 testis and ovary, we performed ATAC-seq on *Drosophila* testis, and obtained ATAC-seq
203 datasets for *Drosophila* Ovarian Somatic Cells (Iwasaki et al., 2016), and S2 cells (Vaid et al.,
204 2020)**.** We annotated peaks in the promoters of genes from each age group, and compared the
205 proportion of genes with detectable high-quality peaks in each tissue (Figure 4A). In every
206 tissue, young genes were the least likely to have detectable chromatin accesibility in their
207 promoters, and old genes were the most likely to have detectable peaks. Every age group of
208 genes was more likely to have peaks in testis, and least likely to have peaks in ovary, indicating
209 that chromatin at the promoter is more broadly open in the testis. In addition, a majority of genes
210 from each age group exhibited more frequent detectable open promoter chromatin in testis. In
211 ovary, by contrast, pre-Bilateria genes are the only age group of which a majority of genes (68%)
212 have detectable ATAC-seq peaks. Every other age group of genes is less likely to contain
213 detectable promoter ATAC-seq peaks, especially young genes, of which only 26% have open
214 chromatin in ovary, compared to 56% of young genes in testis. Our observation that every gene-

215 age group is more likely to have testis peaks than ovary peaks indicates that open chromatin does
216 not underlie the ovary-bias of older genes.
217 The presence of an ATAC-seq peak generally corresponds to increased gene expression
218 in analogous tissues (Figure 4B, 4C, and 4D). Similarly, genes with an ATAC-seq peak in a
219 tissue have heightened activity of their partner TFs compared to genes with no peak in a tissue
220 (Figure 4E, 4F). This indicates that TF expression and promoter chromatin state are useful
221 proxies of a gene's network activity (Sigalova et al., 2020).
222 The low proportion of young genes with ovary ATAC-seq peaks does not entirely explain
223 the paucity of young ovary-biased genes. In the ovary, we found that 13% of young genes are
224 expressed while 26% of them have open promoter chromatin. We therefore sought next to
225 separate the relative influences of TF expression and promoter chromatin for testis and ovary
226 expression
227
228 **High upstream TF expression boosts gene expression in ovary more than in testis**
229 We quantified expression for genes with and without detectable ATAC-seq peaks,
230 conditioning on whether they had high or low TF expression in the tissue (Figure 5). Many genes
231 in testis have surprisingly high expression (median FPKM 1.04) without nearby detectable
232 ATAC-seq peaks or high TF expression, indicating that baseline transcription is higher in testis
233 than in ovary (median FPKM 0.13). Without the aid of many TF partners or open promoter
234 chromatin detectable by ATAC-seq, plenty of genes have surprisingly high expression in testis,
235 but not ovary. In both testis and ovary, the presence of detectable ATAC-seq peaks or high TF
236 expression (greater than the tissue median) is associated with an expression boost. In testis,
237 however, these fold differences in median expression are smaller than in ovary (Table 1).
238 Furthermore, in ovary, high TF expression boosts expression 169.23 fold in genes without a
239 detectable ATAC-seq peak. For genes with a detectable ATAC-seq peak in ovary, high TF
240 expression is associated with a further 15% boost in expression. Ovary expression is 24.18 fold
241 higher for genes with high TF expression but no detectable ATAC-seq peaks compared to genes
242 with open chromatin but low TF expression, indicating that high TF expression is more
243 predictive of expression than chromatin environment in ovary.
244 In testis, both the presence of open chromatin and high TF expression are associated with
245 an expression boost, but every pairwise comparison shows a smaller magnitude difference than

246 in ovary, indicating that trans regulation influences expression in ovary more than testis. While
247 the presence of ATAC-seq peaks correlates with gene expression, TF expression is generally
248 both neccesary and sufficient for gene expression in ovary. Most genes in testis have low but
249 genuine expression (FPKM>1) without nearby detectable ATAC-seq peaks or high TF
250 expression, indicating that leaky transcription may be commonplace in testis. The same category
251 of genes in ovary has a median FPKM of 0.13, negligible by comparison.
252
253 **Discussion:**
254 Our results shape the contours of a model where gene age correlates with tissue-specific
255 determinants of gene expression patterns. Genes are typically born under a simpler regulatory
256 machinery (cis-regulation with fewer TF binding sites (Zhao et al., 2014)), sufficient to drive
257 expression in the testis but not other tissues. As a gene ages, it will likely recruit more trans-
258 acting TF partners, strengthen existing cis-acting TF binding sites (Tuğrul et al., 2015), or gain
259 novel binding sites (Levran et al., 2020; Trizzino et al., 2017). In ovary, the presence of ATAC-
260 seq peaks alone does not correlate with increased expression without help of trans-acting
261 members of a gene's network. The accumulation of TF partners boosts expression in other
262 tissues more than testis, lowering the probability that a middle-aged gene will be testis-biased. Of
263 course, many TF partners are repressive, meaning that their expression would be anti-correlated
264 with that of their target gene. Despite this, older genes with larger TF networks are expressed
265 across a greater variety of tissues, and with consistently higher expression levels than younger
266 genes. Old genes likely continue to recruit more TF sites and relationships, and complex
267 regulatory machinery such as enhancers or insulators. These features only marginally increase
268 expression in the permissive transcriptional environment of the testis, but will substantially
269 increase expression in other tissues, especially ovary. This trend will lead to common ovary-bias
270 of older genes. The age-related complexity of a gene's TF network may therefore drive
271 functional recruitment of young genes to the testis, and old genes to the ovary.
272 DroID does not show whether a TF-gene relationship is one of an activator or repressor.
273 For each gene, we use the same set of TF interactions across every tissue, so the
274 activator/repressor balance should not bias the expression/upstream TF expression relationship
275 between tissues. While our TF expression measure does not consider whether a TF is an activator
276 or repressor, it is still quite predictive of expression. Indeed, the fact that this measure shows a

277 fairly robust Pearson's r with gene expression across tissues might indicate that most of these
278 relationships are activation, consistent with Zhang and Zhou's finding that genes accrue
279 activating TFs and activating epigenetic marks concurrently as they age (Zhang and Zhou, 2019).
280 Our TF expression metric relies on the assumption that most TFs are not an activator of a gene in
281 one tissue and a repressor of the same gene in another tissue. It does not require the assumption
282 that all TF-gene interactions are activation. While TF expression would be a poor method to
283 predict gene expression, TF expression is a useful method to compare the relationship between
284 trans regulation and gene expression between tissues.

285 It is also true that younger gene are often less studied compared to older genes.
286 Fortunately, the DroID TF-gene interaction database shows ChiP-chip profiles of known TFs,
287 giving us a whole-genome holistic comparison of confirmed TF-gene interactions without regard
288 to the age of the target gene. This means that if a TF binds to the promoter of a younger gene, we
289 will still be able to experimentally confirm this interaction even if the gene's function is
290 unknown. This high-throughput approach means that while we have a comprehensive list of
291 confirmed TF-gene interactions, the activation/repression relationship and network modularity of
292 many of these interactions is not yet known barring future lower-throughput experiments.

293 It has been proposed that the testis is uniquely positioned to drive the evolution of new
294 genes due to an open chromatin environment (Assis, 2019; Kaessmann, 2010). Our findings
295 indicate that this general pattern of open chromatin may be reflected on local levels, where we
296 find a substantial proportions of genes of all age groups with ATAC-seq peaks in their
297 promoters. Our findings indicate, however, that TF expression is more predictive of ovary gene
298 expression than the presence of an ATAC-seq peak. This indicates that trans-regulation is
299 especially important for ovarian gene expression, moreso than for testis expression.

300 Even though TF expression is higher in ovary than testis for every age group of genes,
301 this activity does not result in ovary bias for young and middle-aged genes. A fitting analogy is
302 that testis gene expression is like a bicycle in a low gear: easy to initiate movement, but total
303 speed is limited despite the rider's best efforts. Ovary gene expression is more like a bike in high
304 gear: hard to initiate, but given a favorable environment (like biking downhill) the rider can
305 reach greater speeds as a function of their energy input. This may explain why in long-term,
306 ovary becomes a top niche for older genes.

307 Since a good number of the genes in this study originated before multicellular organisms
308 (and therefore animal tissues such as testis and ovary), it is intriguing that such genes are
309 affected by the relationship between gene age and tissue-specificity. Our results do not mean that
310 the fate of all genes is to evolve in testis and gain expression in the ovary. Our results are a
311 snapshot of the relationship between gene age and expression pattern as it occurs now, not a
312 reconstruction of a guaranteed path for the evolution of a given gene's expression.
313 It is instead clear that properties related to gene age differentially influence a gene's
314 potential roles in various tissues. Young genes have relatively few TF binding sites, a state not
315 conducive to expression in most tissues except the testes. Older genes accumulate more TF
316 binding sites (Tuğrul et al., 2015) and gain expression in non-testis tissues. Eventually, adding
317 TF binding sites yields diminishing returns as a gene approaches expression saturation in a
318 tissue. In ovary, however, added TF activity boosts expression more than in other tissues, making
319 ovary-biased expression more likely for older genes. In testis, by comparison, adding TF binding
320 sites appears to have a marginal effect on expression.
321 Future work could focus on the transcription factor aspect of this model. Given that old
322 genes have more TFs than young genes, we would aim to simulate the evolution of a gene's
323 expression trajectory by adding a variable number of TF sites to the promoter of a reporter
324 construct, and analyzing the tissue-specific expression patterns of the construct. This could tell
325 us about the probable evolutionary "fate" of a stereotypical gene's expression: to originate with
326 testis-bias, gain expression in every other tissue, but end with highest expression in ovary. Why
327 ovary currently becomes the top niche remains enigmatic and warrants future studies.
328 Of course, gene expression evolution takes place over millions or billions of years.
329 Newly-originated genes, if they reach fixation in the population, will likely acquire TF sites over
330 time. In another billion years, the regulatory characteristics that today confer testis-bias or ovary-
331 bias may confer bias towards other tissues or even tissues that have not yet emerged.
332
333 **Methods**
334 **Processing of FlyAtlas2 RNA-seq data**
335 Fastq files of adult FlyAtlas2 tissues were obtained from EBI under accession number
336 PRJEB22205 and reads were trimmed with Trimmomatic, set to remove the Illumina universal
337 adapter. Reads were aligned with Hisat2 (Kim et al., 2016), default parameters to the Flybase

338 dmel-r6.15 genome assembly (Thurmond et al., 2019). Reads with mapping quality less than 10
339 were removed. FPKM values were calculated with Stringtie (Kim et al., 2016) using default
340 parameters. For each gene, FPKMs were averaged across replicates of a tissue.
341
342 **Determination of consensus gene ages**
343 To allow for better statistical power and relatively uniform group sizes between gene age
344 groups we binned genes into 3 groups: genes that emerged after the pan-Drosophilid divergence
345 (Drosophilid), genes that emerged sometime before the pan-Drosophilid divergence but before
346 the divergence of Bilateria (pre-Drosophilid), and genes that emerged before Bilateria (pre-
347 Bilateria). To define Drosophilid genes, we used genes assigned to branches 1-5 in the gene age
348 dataset from Zhang et al. (Zhang et al., 2010b). Ages of older genes were assigned using gene
349 ages from Kondo et al (Kondo et al., 2017). Genes without ages defined in either dataset were
350 not included for figures that segment genes by age, but were included for analyses of TF
351 expression and open chromatin that did not consider gene age. For supplemental figures we
352 reproduced the main figures defining genes from all 3 age groups only according to the ages
353 assigned by Kondo et al. (Kondo et al., 2017) and observed no differences that would change our
354 main findings.
355
356 **Calculation of tissue specificity**
357 We used the tau method (Kryuchkova-Mostacci and Robinson-Rechavi, 2017) to
358 calculate tissue specificity based on a gene's FPKM across adult tissues, with replicates averaged
359 (Kryuchkova-Mostacci and Robinson-Rechavi, 2017). A tau close to 1 indicates a tissue-specific
360 gene, with a tau of 1 indicating a gene is only expressed in one tissue. A tau close to zero
361 indicates a gene is equally expressed in every tissue.
362
363 **Calculation of scaled gene expression**
364 FPKM is not normalized between genes, so we scaled gene expression to compare genes
365 with different thresholds of activity. For a tissue i, scaled expression of a gene j is log-
366 transformed FPKM in tissue i divided by gene j's max logFPKM in any tissue. A scaled
367 expression of 1 is gene j's maximum expression in any tissue, and a scaled expression of 0

368 means expression is not detected. A scaled expression of 0.5 means that the logFPKM of a gene
369 in a particular tissue is half the maximum observed logFPKM in any tissue.
370
371 **Calculation of TF expression for genes/tissues**
372 For each gene, we wanted a measure for the activity of its upstream regulators in every
373 tissue. We used the DroID database (Murali et al., 2011), which, for over 700 TFs, lists all genes
374 whose promoters are bound by each TF as annotated with ChIP-chip and ChIP-seq by the
375 modENCODE project (Roy et al., 2010). For this analysis, we only used genes with at least one
376 TF annotated by DroID.
377 For a gene in a tissue, the TF expression score is the summed scaled expression of all
378 annotated TF partners of that gene in that tissue. For example: if a gene's TF partners have
379 scaled expression values of 1, 1, and 0.5 in a tissue, and 0, 0, 0.5 in another tissue, the activity
380 score for that gene would be 2.5 in the first tissue and 0.5 in the second, reflecting higher
381 network activity in the first tissue. Since the TF expression values are scaled first, this measure
382 allows for holistic comparisons of TF expression patterns between genes and tissues. The
383 correlation between open promoter chromatin and TF expression in multiple tissues assures us
384 that this metric measures biologically meaningful activity.
385
386 **ATAC-seq of *Drosophila* testis**
387 We performed ATAC-seq experiment and analysis using 2-day-old testis of *D.*
388 *melanogaster* RAL517 stain. For each sample, 25 newly emerged males were collected and
389 transferred to 3 new vials (performed in triplicate). 48 hours later, we dissected testes in cold
390 PBS. Tissues were lysed in 200 µl of ATAC-Seq lysis buffer (10mM Tris-HCl, 10mM NaCl,
391 3mM $MgCl_2$, 0.1% IGEPAL CA-630) and manually homogenized with a plastic pestle, followed
392 by a 1-minute incubation on ice, this process was repeated three times. The samples were
393 pelleted at 4°C (100g for 10 minutes) to recover the nuclei. The buffer was removed and the
394 nuclear pellet was re-suspended in 200 µl of lysis buffer. The nuclei preparation was filtered
395 through a 30 µm Nitex nylon mesh (Genesee Scientific #57-105); the filter was further washed
396 with another 200 µl of lysis buffer to ensure optimal nuclear recovery. The purified nuclei were
397 isolated by centrifugation at 1000g for 10 minutes at 4°C). Following buffer removal, the nuclei
398 were processed for the tagmentation reaction by adding: 12.5 µl Nextera Tagment DNA Buffer,

399 11.25 µl ddH2O and 1.25 µl Tn5 Transposase (Illumina Kit # FC-121-1030). The reaction was
400 carried out in a thermal cycler for 30 minutes at 37°C with an additional mixing step 15 minutes
401 into the reaction. The fragments were then purified using the Qiagen MinElute PCR purification
402 kit (#28004) according to instructions. Libraries were constructed using the same primers as
403 Buenrostro et al. (Buenrostro et al., 2015) and following a similar workflow: the purified DNA
404 was first amplified for 5 cycles by PCR using the NEB Ultra II PCR mix (M0544). Then, an
405 aliquot of the PCR reaction was analyzed by qPCR to determine the remaining optimal number
406 of PCR cycles. Libraries were finally purified using SPRI beads with a two-step size selection
407 protocol with bead-to-sample ratios of 0.55× and 1.00× for the first and second step,
408 respectively. An aliquot of the purified library was used for quality control, and tested on an
409 Agilent D1000 Tapestation platform, where concentration and peak periodicity were assessed.
410 The samples were additionally tested for quality using Qubit, and sequenced on a 75bp paired-
411 end Hiseq X platform.

412

413 **Processing of ATAC-seq data from testis, ovary, and S2 cells**

414 We generated 3 replicates of testis ATAC-seq data from *D. melanogaster*. 2 replicates of
415 OSC data were used: SRR3503078 and SRR3503086 (Iwasaki et al., 2016). 2 replicates of S2
416 cells were used: SRR5985082 and SRR5985083 (Ibrahim et al., 2018). Reads were aligned with
417 bowtie2 (Langmead and Salzberg, 2012), default parameters against the flybase dmel_r6.24
418 reference genome (Thurmond et al., 2019). BAM files for each tissue were then merged with
419 samtools merge (Li et al., 2009). Macs2 (Zhang et al., 2008) was used to call peaks for each
420 tissue with the –nomodel parameter. The narrowpeak files were then loaded into R for further
421 processing with Chippeakanno (Zhu et al., 2010) (details in supplementary Rmd on Github).
422 Only peaks with a q value < 0.05 were used. Chippeakanno was run to find peaks overlapping
423 the region 2000 bp upstream – 100 bp downstream of every gene's TSS.

424

425 **Data availability**

426 Scripts and processed data needed to reproduce figures are deposited in
427 https://github.com/LiZhaoLab/TissueSpecificity. Testis ATAC-seq of *Drosophila melanogaster*
428 Ral517 is deposited at NCBI under biosample accession # SAMN16259271.

429

Data reproducibility

The data needed to reproduce this work can be found in this link https://github.com/LiZhaoLab/TissueSpecificity. It includes calculated FPKM for every gene and tissue in FlyAtlas2, files used to calculate consensus gene ages from Kondo et al. and Zhang et al., narrowpeak files we calculated for each of the 3 ATAC-seq datasets, a csv file with calculated TF expression (connectivity.csv) for each gene and tissue, and a file from DroID showing every experimentally annotated TF-gene interaction (tf_gene.txt). These files are all referenced by the Rmd script on our Github page. The free web app which allows users to interactively explore our tissue specificity data for any set of genes without coding experience necessary is: https://zhao.labapps.rockefeller.edu/tissue-specificity/.

Acknowledgements

We are grateful to Stuart McDonald from The University of Kansas and Margo Herre from Leslie Vosshall's lab at The Rockefeller University for the help and suggestions on ATAC-seq, and Luis Gracia at The Rockefeller University for the help with the web app. We thank the members of the Zhao lab for helpful discussions and critically reading an earlier version of the manuscript.

Author contribution

E.W. and L.Z. conceived the study. N.S. and S.B. generated the ATAC-seq data. E.W. performed all the analysis in this manuscript. E.W. and L.Z. wrote the manuscript with the input from all authors.

Funding

The work was supported by NIH MIRA R35GM133780. L.Z. was supported by the Robertson Foundation, a Monique Weill-Caulier Career Scientist Award, an Alfred P. Sloan Research Fellowship (FG-2018-10627), a Rita Allen Foundation Scholar Program, and a Vallee Scholar Program (VS-2020-35).

Declaration of interests

The authors declare no competing interests.

461 **References**

462 Assis, R. (2019). Out of the testis, into the ovary: biased outcomes of gene duplication and
463 deletion in Drosophila. Evolution (N. Y). *73*, 1850–1862.
464 Begun, D.J., Lindfors, H.A., Thompson, M.E., and Holloway, A.K. (2006). Recently Evolved
465 Genes Identified From Drosophila yakuba and D. erecta Accessory Gland Expressed Sequence
466 Tags. Genetics *172*, 1675–1681.
467 Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A method for
468 assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *2015*, 21.29.1-21.29.9.
469 Ding, Y., Zhao, L., Yang, S., Jiang, Y., Chen, Y., Zhao, R., Zhang, Y., Zhang, G., Dong, Y., Yu,
470 H., et al. (2010). A young Drosophila duplicate gene plays essential roles in spermatogenesis by
471 regulating several Y-linked male fertility genes. PLoS Genet. *6*, e1001255.
472 Grive, K.J., Hu, Y., Shu, E., Grimson, A., Elemento, O., Grenier, J.K., and Cohen, P.E. (2019).
473 Dynamic transcriptome profiles within spermatogonial and spermatocyte populations during
474 postnatal testis maturation revealed by single-cell sequencing. PLOS Genet. *15*, e1007810.
475 Ibrahim, M.M., Karabacak, A., Glahs, A., Kolundzic, E., Hirsekorn, A., Carda, A., Tursun, B.,
476 Zinzen, R.P., Lacadie, S.A., and Ohler, U. (2018). Determinants of promoter and enhancer
477 transcription directionality in metazoans. Nat. Commun. *9*, 4472.
478 Iwasaki, Y.W., Murano, K., Ishizu, H., Shibuya, A., Iyoda, Y., Siomi, M.C., Siomi, H., and
479 Saito, K. (2016). Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors
480 Including Histone H1 to Repress Transposons. Mol. Cell *63*, 408–419.
481 Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. Genome Res.
482 *20*, 1313–1326.
483 Kim, D., Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-
484 level expression analysis of RNA- seq experiments with HISAT , StringTie and Transcript-level
485 expression analysis of RNA-seq experiments with HISAT , StringTie and Ballgown. Nat. Protoc.
486 *11*, 1650–1667.
487 Kondo, S., Vedanayagam, J., Mohammed, J., Eizadshenass, S., Kan, L., Pang, N., Aradhya, R.,
488 Siepel, A., Steinhauer, J., and Lai, E.C. (2017). New genes often acquire male- specific functions
489 but rarely become essential in Drosophila. 1841–1846.
490 Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression
491 tissue-specificity metrics. Brief. Bioinform. *18*, 205–214.

492 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat.
493 Methods *9*, 357–359.
494 Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A., and Dow, J.A.T. (2018). FlyAtlas 2: A new
495 version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-
496 specific data. Nucleic Acids Res. *46*, D809–D815.
497 Levine, M.T., Jones, C.D., Kern, A.D., Lindfors, H.A., and Begun, D.J. (2006). Novel genes
498 derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit
499 testis-biased expression. Proc. Natl. Acad. Sci. *103*, 9935–9939.
500 Levran, O., Even-Tov, E., and Zhao, L. (2020). A hominid-specific shift in cerebellar expression,
501 upstream retrotransposons, and a potential cis-regulatory mechanism: bioinformatics analyses of
502 the mu-opioid receptor gene. Heredity (Edinb). *124*, 325–335.
503 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
504 Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence
505 Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
506 Long, M., Betrán, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses
507 from the young and old. Nat. Rev. Genet. *4*, 865–875.
508 Long, M., VanKuren, N.W., Chen, S., and Vibranovski, M.D. (2013). New Gene Evolution:
509 Little Did We Know. Annu. Rev. Genet. *47*, 307–333.
510 Majic, P., and Payne, J.L. (2020). Enhancers facilitate the birth of de novo genes and gene
511 integration into regulatory networks. Mol. Biol. Evol. *37*, 1165–1178.
512 Mayr, C. (2016). Evolution and Biological Roles of Alternative 3'UTRs. Trends Cell Biol. *26*,
513 227–237.
514 Murali, T., Pacifico, S., Yu, J., Guest, S., Roberts, G.G., and Finley, R.L. (2011). DroID 2011: A
515 comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions
516 for Drosophila. Nucleic Acids Res. *39*, D736–D743.
517 Ohno, S. (1970). Evolution by Gene Duplication (Berlin, Heidelberg: Springer Berlin
518 Heidelberg).
519 Palmieri, N., Kosiol, C., and Schlötterer, C. (2014). The life cycle of Drosophila orphan genes.
520 Elife *3*, e01311.
521 Roy, S., Ernst, J., Kharchenko, P. V, Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M.,
522 Bristow, C.A., Ma, L., Lin, M.F., et al. (2010). Identification of functional elements and

523 regulatory circuits by Drosophila modENCODE. Science *330*, 1787–1797.
524 Schmidt, E.E. (1996). Transcriptional promiscuity in testes. Curr. Biol. *6*, 768–769.
525 Sigalova, O.M., Shaeiri, A., Forneris, M., Furlong, E.E.M., and Zaugg, J.B. (2020). Predictive
526 features of gene expression variation reveal mechanistic link with differential expression. Mol.
527 Syst. Biol. *16*, e9539.
528 Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P.,
529 Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular Source and Mechanisms of High
530 Transcriptome Complexity in the Mammalian Testis. Cell Rep. *3*, 2179–2190.
531 Tautz, D., and Domazet-Lošo, T. (2011). The evolutionary origin of orphan genes. Nat. Rev.
532 Genet. *12*, 692–702.
533 Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J.,
534 Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al. (2019). FlyBase 2.0: the next
535 generation. Nucleic Acids Res. *47*, D759–D765.
536 Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry,
537 G.H., Lynch, V.J., and Brown, C.D. (2017). Transposable elements are the primary source of
538 novelty in primate gene regulation. Genome Res. *27*, 1623–1633.
539 Tuğrul, M., Paixão, T., Barton, N.H., and Tkačik, G. (2015). Dynamics of Transcription Factor
540 Binding Site Evolution. PLoS Genet. *11*, e1005639.
541 Vaid, R., Wen, J., and Mannervik, M. (2020). Release of promoter-proximal paused Pol II in
542 response to histone deacetylase inhibition. Nucleic Acids Res. *48*, 4877–4890.
543 Vinckenbosch, N., Dupanloup, I., and Kaessmann, H. (2006). Evolutionary fate of retroposed
544 gene copies in the human genome. Proc. Natl. Acad. Sci. *103*, 3220–3225.
545 Witt, E., Benjamin, S., Svetec, N., and Zhao, L. (2019). Testis single-cell RNA-seq reveals the
546 dynamics of de novo gene transcription and germline mutational bias in Drosophila. Elife *8*,
547 e47138.
548 Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S.Y., Keefe, D.L.,
549 Alukal, J.P., Boeke, J.D., et al. (2020). Widespread Transcriptional Scanning in the Testis
550 Modulates Gene Evolution Rates. Cell *180*, 248-262.e21.
551 Zhang, J.-Y., and Zhou, Q. (2019). On the Regulatory Evolution of New Genes Throughout
552 Their Life History. Mol. Biol. Evol. *36*, 15–27.
553 Zhang, L., Tan, Y., Fan, S., Zhang, X., and Zhang, Z. (2019). Phylostratigraphic analysis of gene

554 co-expression network reveals the evolution of functional modules for ovarian cancer. Sci. Rep.
555 *9*, 2623.
556 Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C.,
557 Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS).
558 Genome Biol. *9*, R137–R137.
559 Zhang, Y.E., Vibranovski, M.D., Landback, P., Marais, G.A.B., and Long, M. (2010a).
560 Chromosomal Redistribution of Male-Biased Genes in Mammalian Evolution with Two Bursts
561 of Gene Gain on the X Chromosome. PLoS Biol. *8*, e1000494.
562 Zhang, Y.E., Vibranovski, M.D., Krinsky, B.H., and Long, M. (2010b). Age-dependent
563 chromosomal distribution of male-biased genes in Drosophila. Genome Res. *20*, 1526–1533.
564 Zhang, Z., Hambuch, T.M., and Parsch, J. (2004). Molecular evolution of sex-biased genes in
565 Drosophila. Mol. Biol. Evol. *21*, 2130–2139.
566 Zhao, L., Saelao, P., Jones, C.D., and Begun, D.J. (2014). Origin and spread of de novo genes in
567 Drosophila melanogaster populations. Science (80-.). *343*, 769–772.
568 Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and
569 Wang, W. (2008). On the origin of new genes in Drosophila. Genome Res. *18*, 1446–1455.
570 Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R.
571 (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC
572 Bioinformatics *11*, 237.
573
574 **Tables and Figure Legends**
575

Category	Ovary median FPKM	Ovary fold difference	Testis median FPKM	Testis fold difference
Low TF expression, no ATAC peak Low TF expression, ATAC peak	0.13 0.91	7.00	1.04 3.50	3.37
Low TF expression, no ATAC peak High TF expression, no ATAC peak	0.13 22.00	169.23	1.04 7.20	6.92
Low TF expression, no ATAC peak High TF expression, ATAC peak	0.13 26.10	200.77	1.02 10.10	9.90
Low TF expression, ATAC peak	0.91	24.18	3.50	2.06

High TF expression, no ATAC peak	22.00		7.20	
Low TF expression, ATAC peak High TF expression, ATAC peak	0.91 26.10	28.68	3.50 10.10	2.89
High TF expression, no ATAC peak High TF expression, ATAC peak	22.00 26.10	1.19	7.20 10.10	1.40

576

577 **Table 1: High TF expression confers a disproportionate fold difference in gene expression in ovary.**
578 Corresponding to the median values shown in figure 5, these are the pairwise fold differences in median FPKM for
579 genes with and without promoter peaks, and genes with upstream TF expression above or below the median for a
580 tissue. In ovary, genes with no detectable peak have 169.23-fold higher expression if their TF expression is higher
581 than the median TF expression for genes in ovary. In testis, fold differences in median expression are much smaller
582 between groups.
583

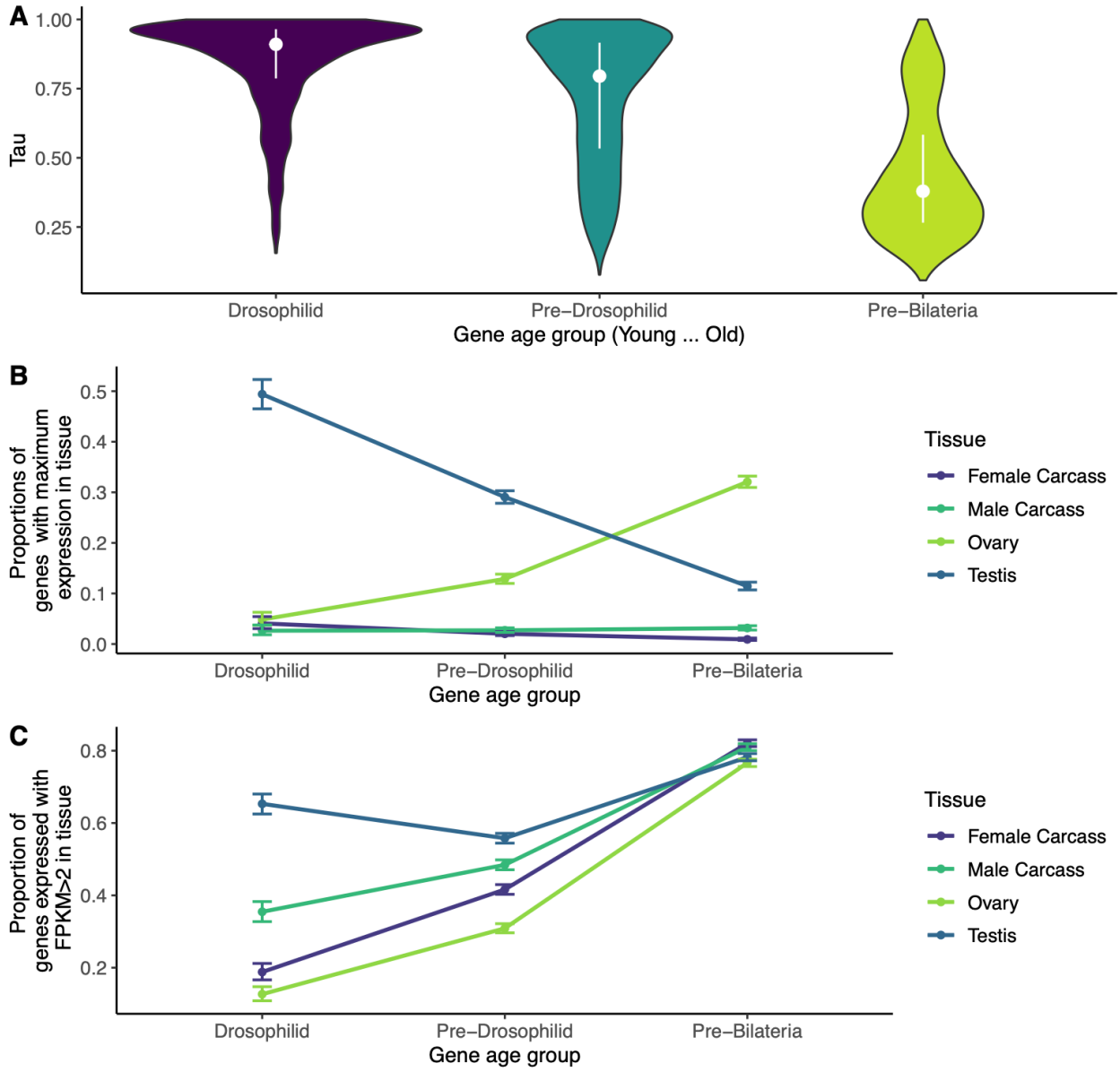


Figure 1: Young genes are testis-specific, old genes are broadly-expressed and often ovary-biased. A) Average Tau values among genes of each age group, dots represent medians and vertical lines are interquartile ranges. Young genes are more tissue specific (higher Tau) than older genes. B) For four tissues, proportion of genes of each age group with maximum expression in that tissue. Younger genes usually have highest expression in testis, but the proportion of testis-biased genes declines with gene age. Ovary biased young genes are rare, but old genes are more often biased towards ovary than any other tissue. Error bars are 95 percent confidence intervals for proportion test. C) For four tissues, proportion of genes of each age group with FPKM >2 in that tissue. In testis, ovary and carcass, old genes are more likely to be expressed than young genes, but this disparity is smallest in testis and largest in ovary. By this measure, old genes are no longer biased in testis. Ovary-bias of older genes is not explained by the relative proportion of genes expressed between tissues.

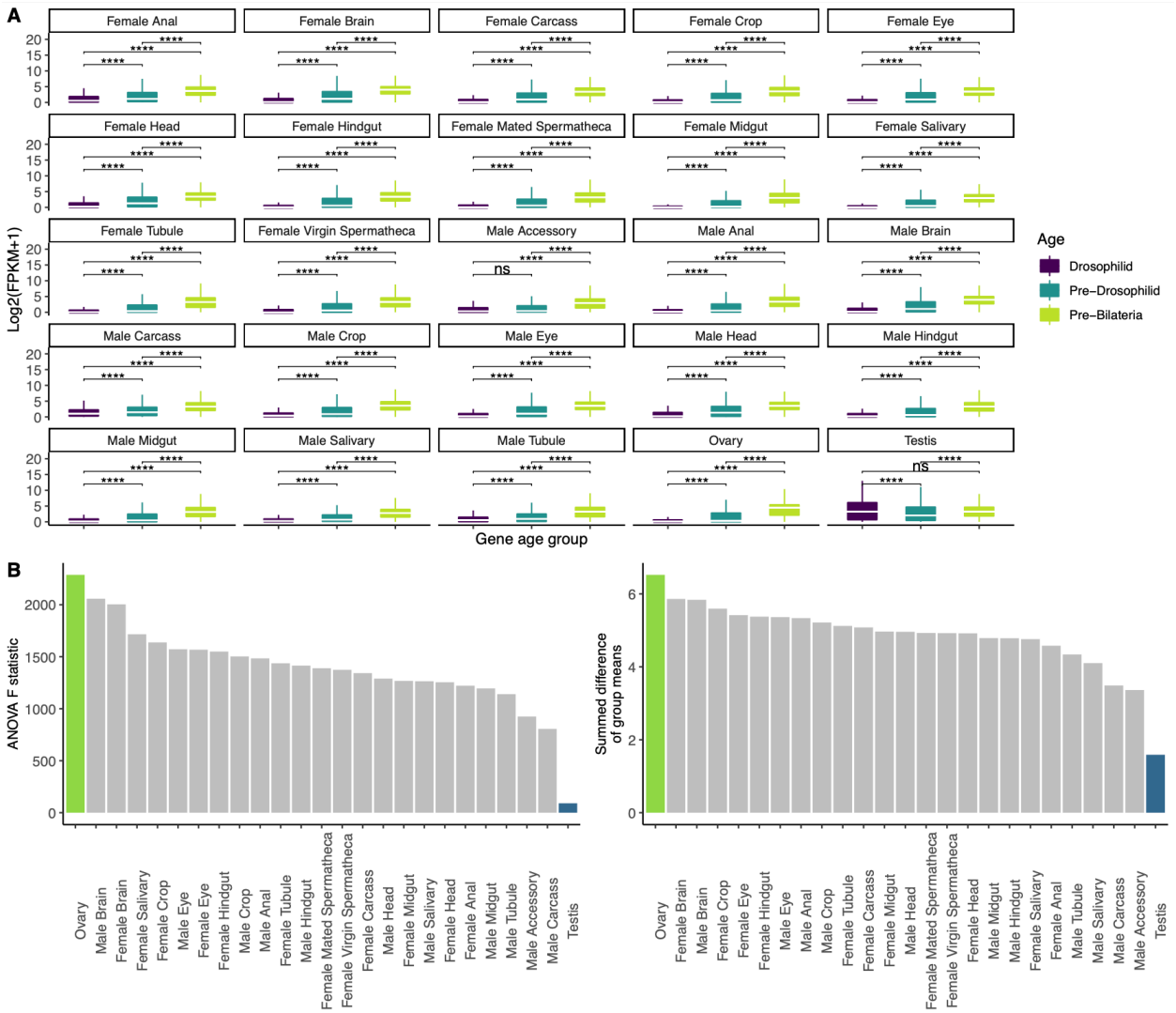


Figure 2: Tissue-specific trends in the gene-age/expression relationship. A) The log-scaled expression of every gene in that tissue versus gene age for every adult tissue (see methods). In almost every tissue scaled expression is very low for young genes, and high for older genes. The testis is an outlier, with statistically similar expression between Drosophilid and Pre-Drosophilid genes. Asterisks represent P values are adjusted with Bonferroni's correction (*=p<0.05, **=p<0.005,***=p<0.0005, ****=p<0.00005). Raw and Bonferroni-adjusted p values are in Supplemental Table 1. B) Rankings of ANOVA statistics for all tissues. We performed an ANOVA on each of the panels from part A, comparing, for each tissue, the ratio of inter-group variation to between group variation (F statistic). By this measure, ovary has the largest relationship between gene age and expression (because old genes are often ovary biased), and testis has the smallest (because old and young genes are similarly expressed in the testis). We also took the mean difference between groups and summed their absolute values for each tissue. Testis has the smallest mean expression difference between age groups, and ovary has the largest. This conclusion held when we repeated the analysis using an alternate set of gene age assignments (Supplemental Figure 6).

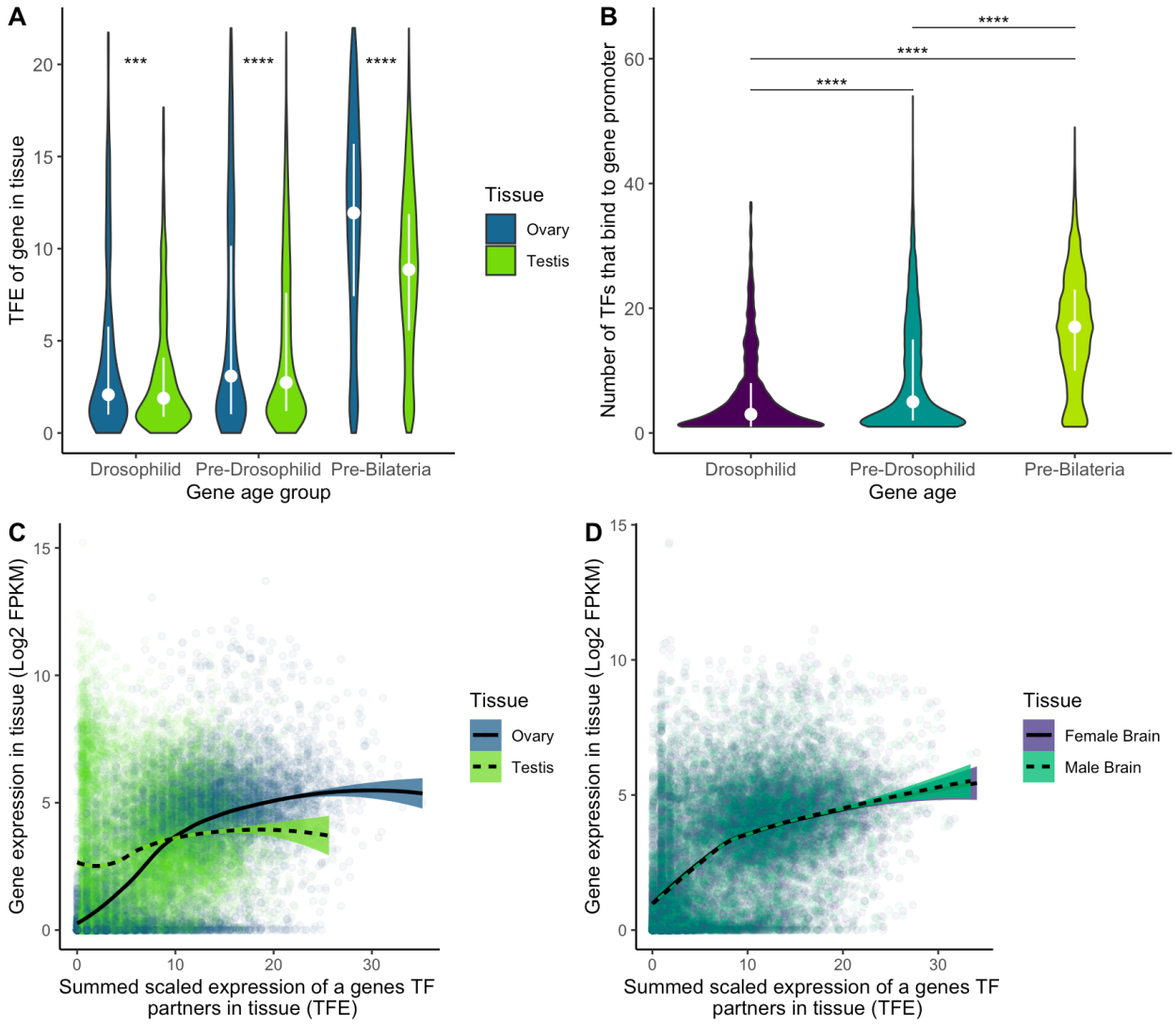


Figure 3: Transcription factor expression explains gene age/expression trends in testis and ovary. A) Upstream TF expression in testis and ovary for genes with different ages, using DroID's curated database of TF binding profiles from the modENCODE project. For every gene with a confirmed TF-promoter interaction, we calculated TF expression in testis and ovary by scaling the expression for each TF from 0 to 1, and summing the scaled expression of every TF that binds to the promoter of a given gene. Older genes have much higher TF expression than younger genes in both tissues, and no age group of genes shows elevated TF expression in testis compared to ovary. White dots are medians and lines are interquartile ranges. Asterisks represent p values adjusted with Bonferroni's correction. B) In DroID data, the promoters of older genes have been shown able to be bound by more TFs than younger genes. C) Log-scaled gene expression versus upstream TF expression in gonads. Genes require less upstream TF expression for expression in testis than in ovary. Genes have fairly high testis expression even without much TF expression in testis, but genes with low ovary TF expression are relatively lowly expressed in ovary. This indicates that genes require less TF expression for testis expression than ovary expression. In ovary, higher TF expression corresponds to higher expression, moreso than testis, where adding TF expression makes

relatively little difference in testis gene expression. D) Log-scaled gene expression versus upstream TF expression in brains. These sexually dimorphic tissues show no difference in their relationships between TF expression and gene expression. In these tissues, low TF expression yields low expression, and high TF expression yields high expression, much like the ovary and much unlike the testis. Lines are smoothed loess regressions with 95 percent confidence intervals. Other tissues are shown in Supplemental Figure 3.

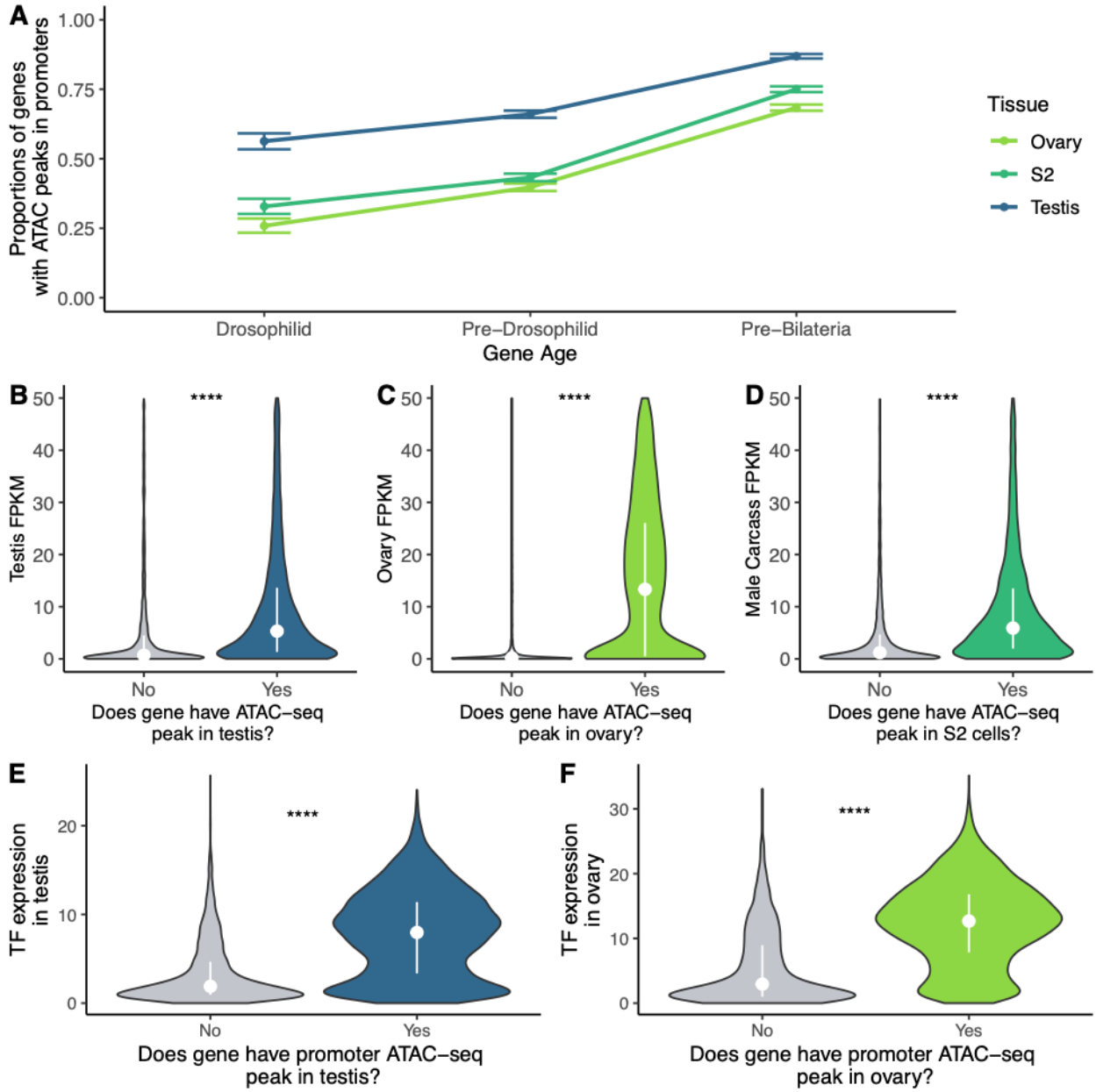
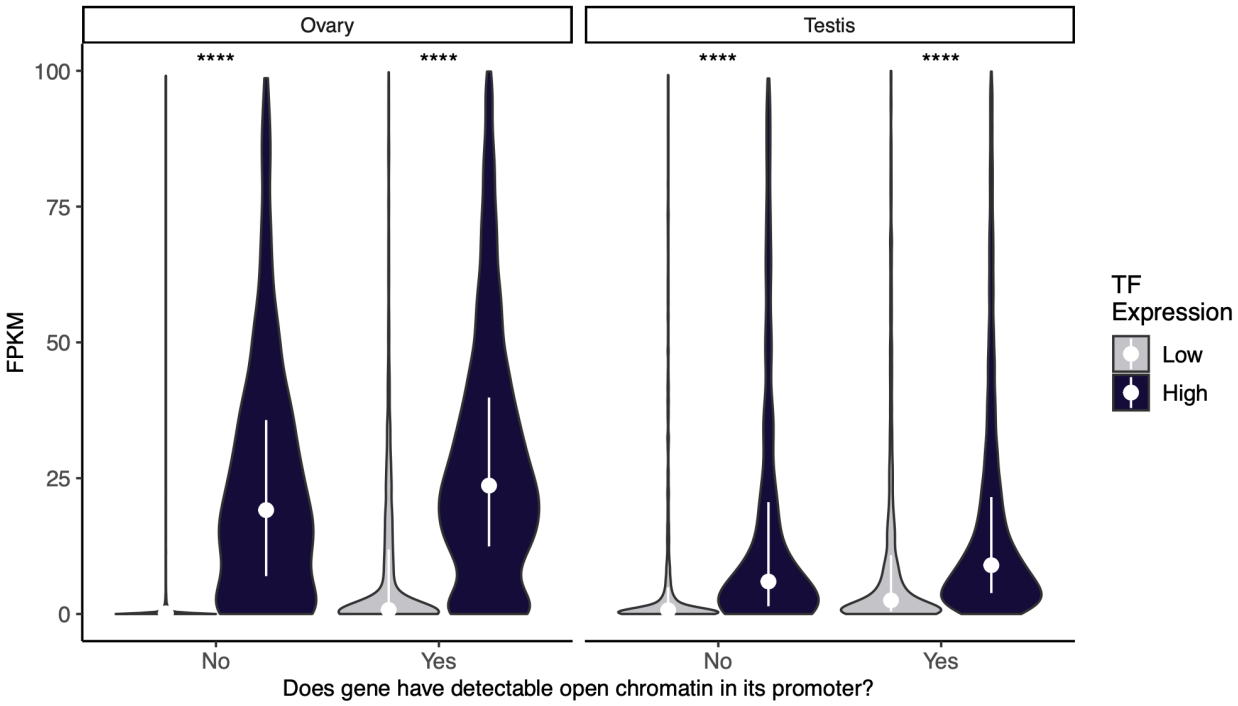


Figure 4: ATAC-seq peaks show an age-related trend in multiple tissues. A) The relative proportions of genes with a detectable ATAC-seq peak in their promoters, for 3 gene age groups and 3 datasets. For each dataset, young genes were the least likely to have open chromatin in their promoters. Testis is unique among these datasets because a majority of genes of each age group have open promoter chromatin. B) FPKM for genes with and without a

636 detected promoter ATAC-seq peak in testis. Genes with open promototer chromatin in testis have generally higher
637 expression in FlyAtlas2 data. Dots are medians, white line is the interquartile range. C) Genes with open promoter
638 chromatin in ovary have higher FlyAtlas2 expression, and the FPKM difference between genes with and without
639 peaks is much larger than the other two tissues. D) Genes with promoter peaks in S2 cells generally have higher
640 expression in male carcass, the most analogous FlyAtlas2 tissue to this cell line. E) TF expression for genes with and
641 without detectable ATAC-seq peaks. Genes with ATAC-seq promoter peaks tend to have higher TF expression in
642 testis, F) as well as ovary. **** represents adjusted p values <0.00005.
643



644

645 **Figure 5: High TF expression disproportionately predicts gene expression in ovary.** For testis and ovary,
646 FPKM for genes with and without detectable chromatin peaks, grouped by "high" or "low" upstream TF expression.
647 Genes are classified as high or low activity in a tissue if they are above or below the median TF expression for genes
648 in the tissue. In testis, both high TF expression and open promoter chromatin confer a similar, modest expression
649 benefit. In ovary, genes with low TF expression are generally very lowly expressed regardless of the presence of a
650 promoter peak. This indicates that TF expression influences ovary expression more than chromatin environment. In
651 ovary, high TF expression is necessary and sufficient for gene expression. White dots are the median values for each
652 group, used to calculate fold changes in table 1. Vertical lines are inter-quartile ranges.

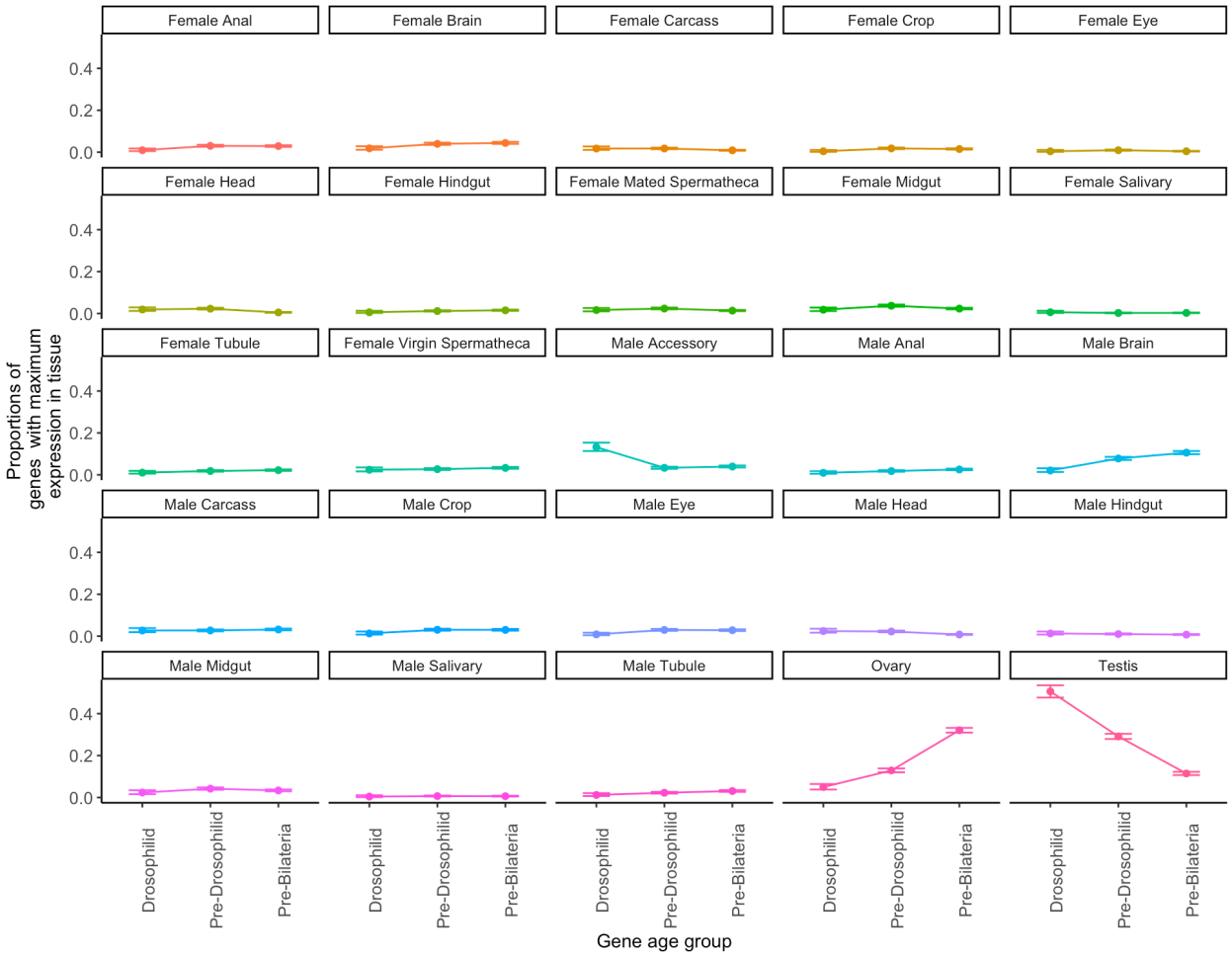
Supplementary figures and tables for

Transcription factors drive opposite relationships between gene age and tissue specificity in male and female *Drosophila* gonads

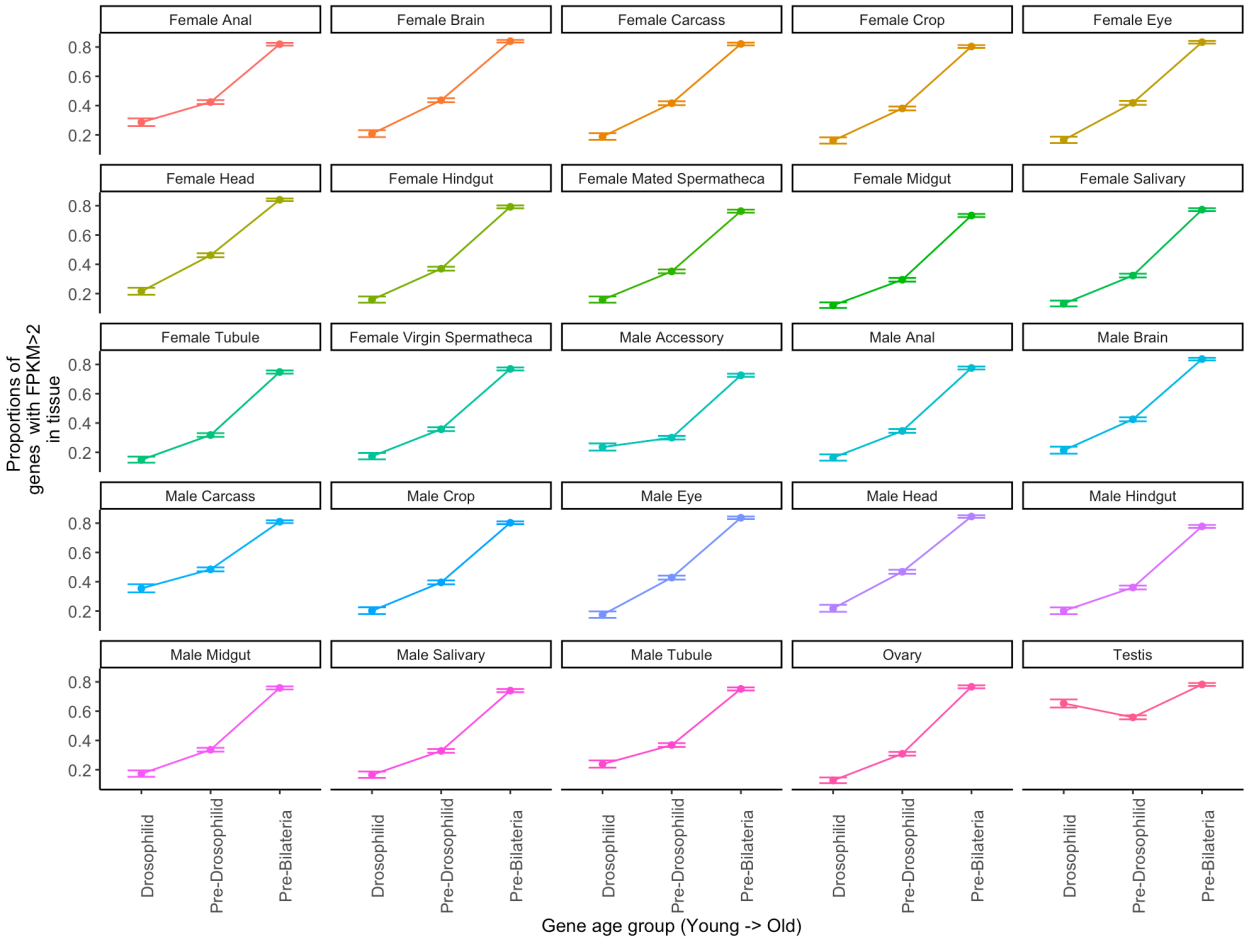
Evan Witt, Nicolas Svetec, Sigi Benjamin, Li Zhao*

Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY 10065, USA

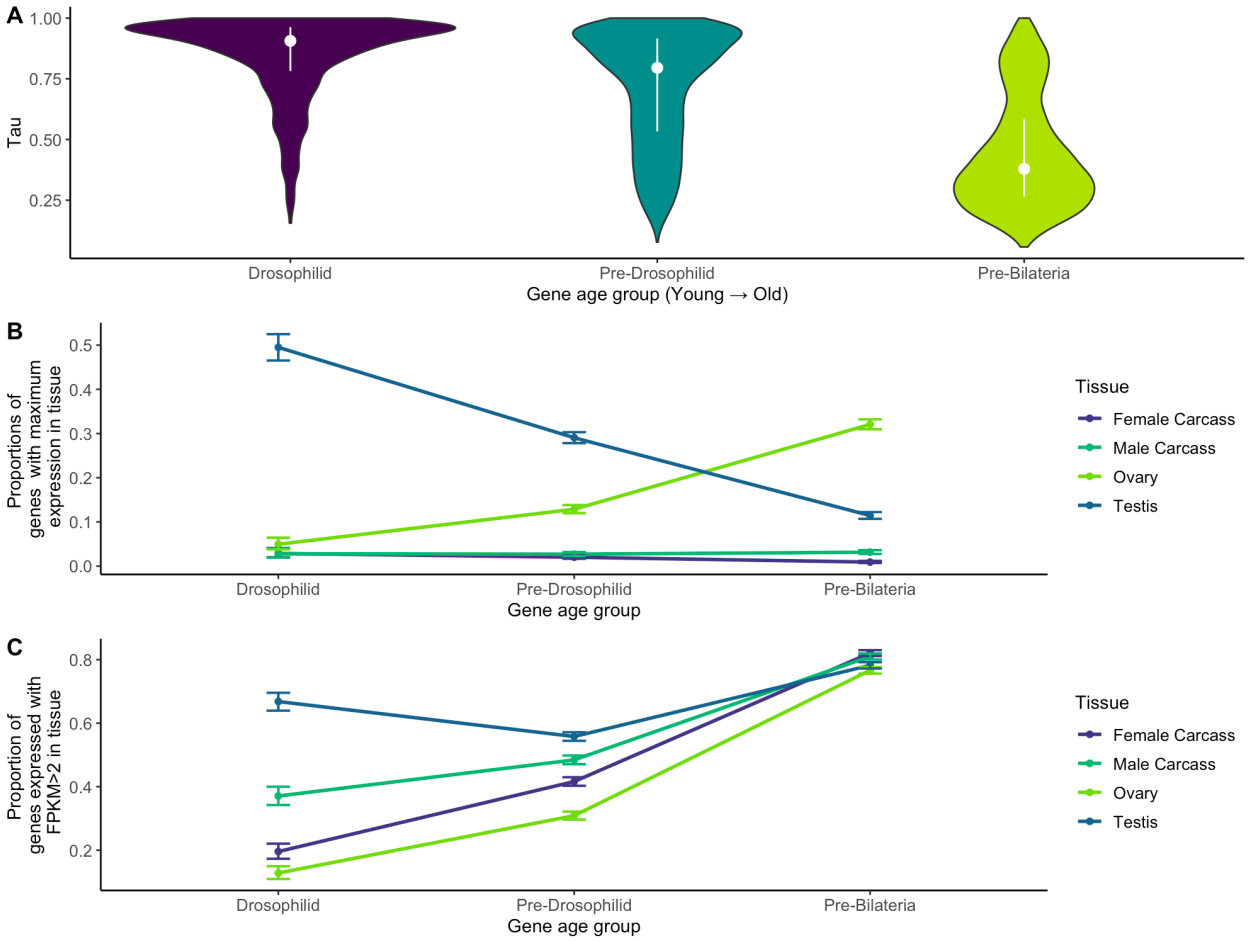
*Correspondence to: lzhao@rockefeller.edu



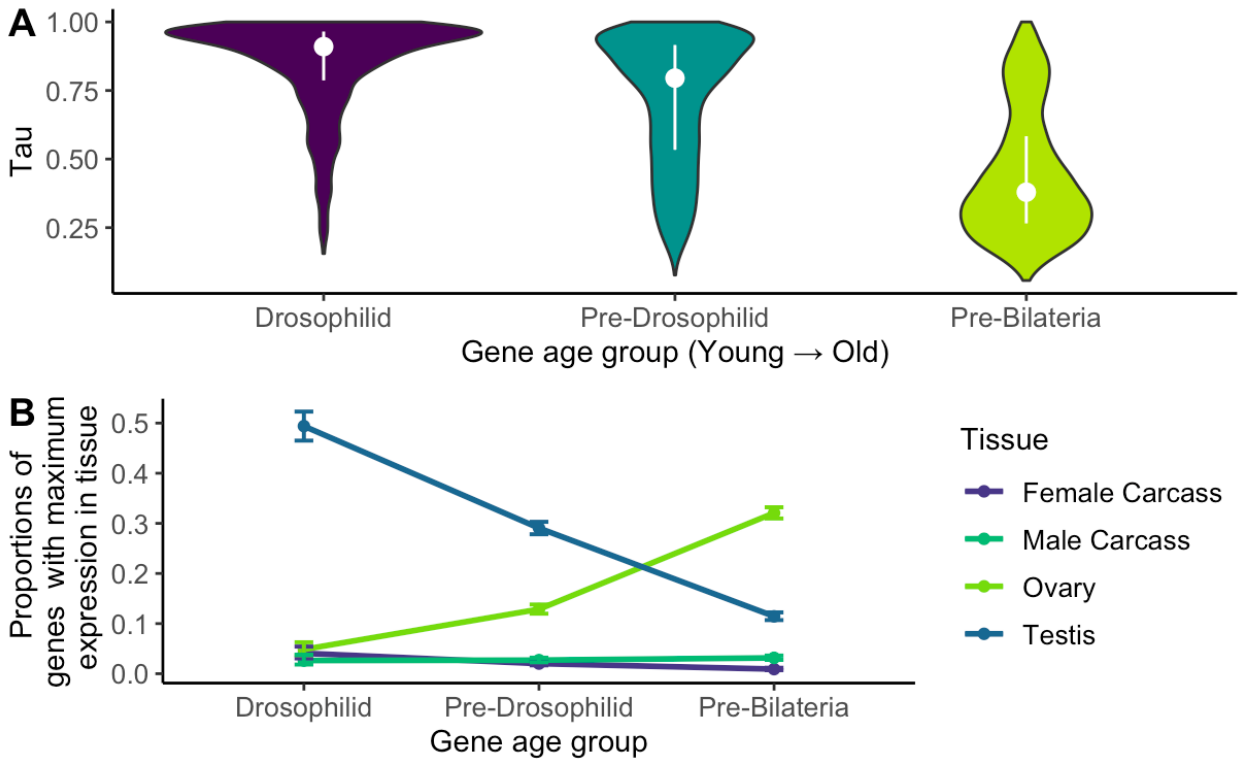
Supplemental Figure 1: Proportion of genes with maximum expression in a tissue, by age group. A high proportion of young genes have maximum expression in testis, and a high proportion of older genes have maximum expression in the ovary. Between the testis and accessory gland, a majority of Drosophilid genes are biased towards the male reproductive system. Some tissues like male brain and male accessory gland show a trend in age-related tissue bias, but none as large as the testis and ovary.



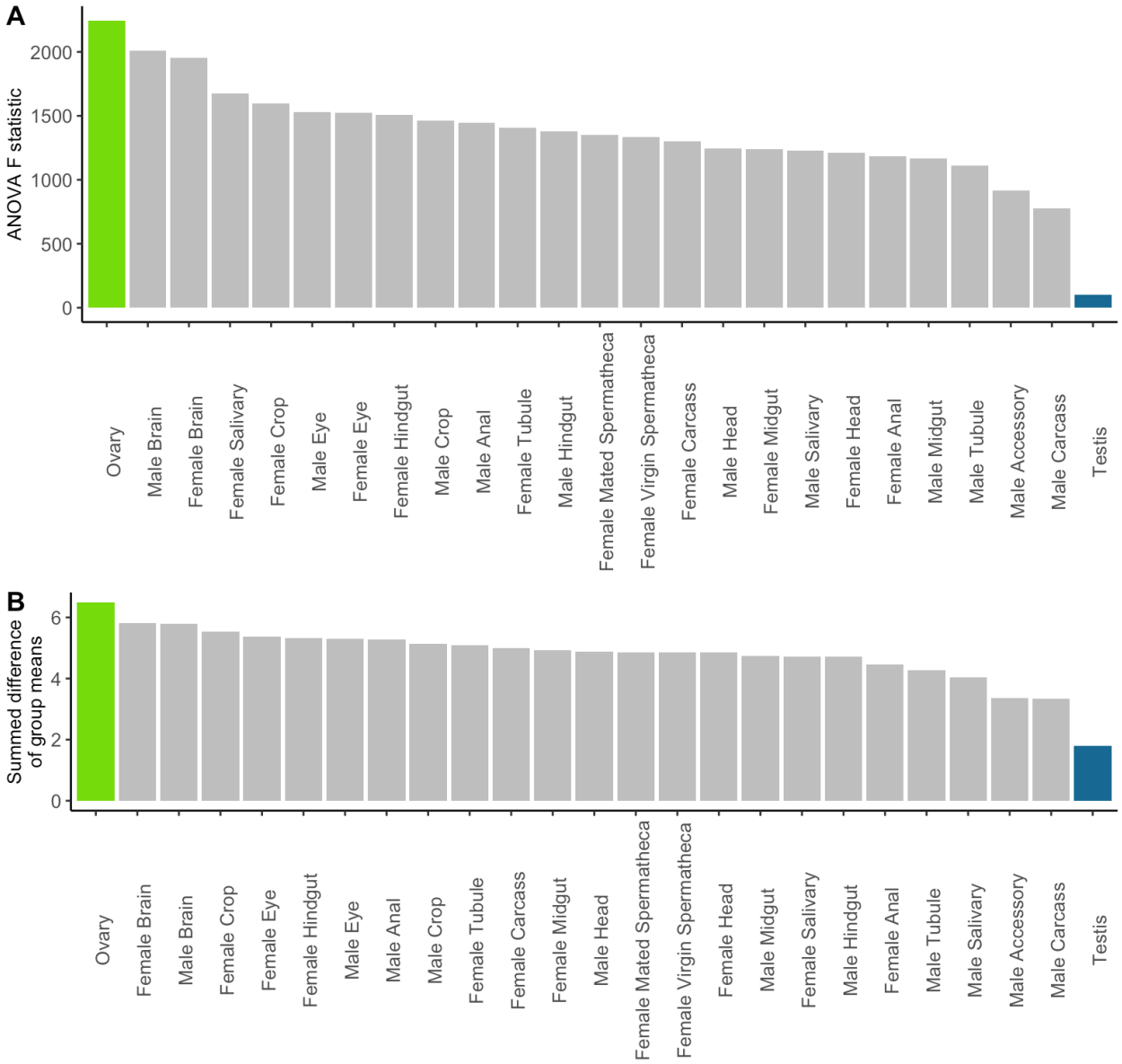
Supplemental Figure 2: Proportion of genes expressed with FPKM > 2, by Fylatlas2 tissue and age. In every tissue, old genes are most commonly expressed, but ovary has the greatest difference between young and old genes, and testis has the smallest difference.



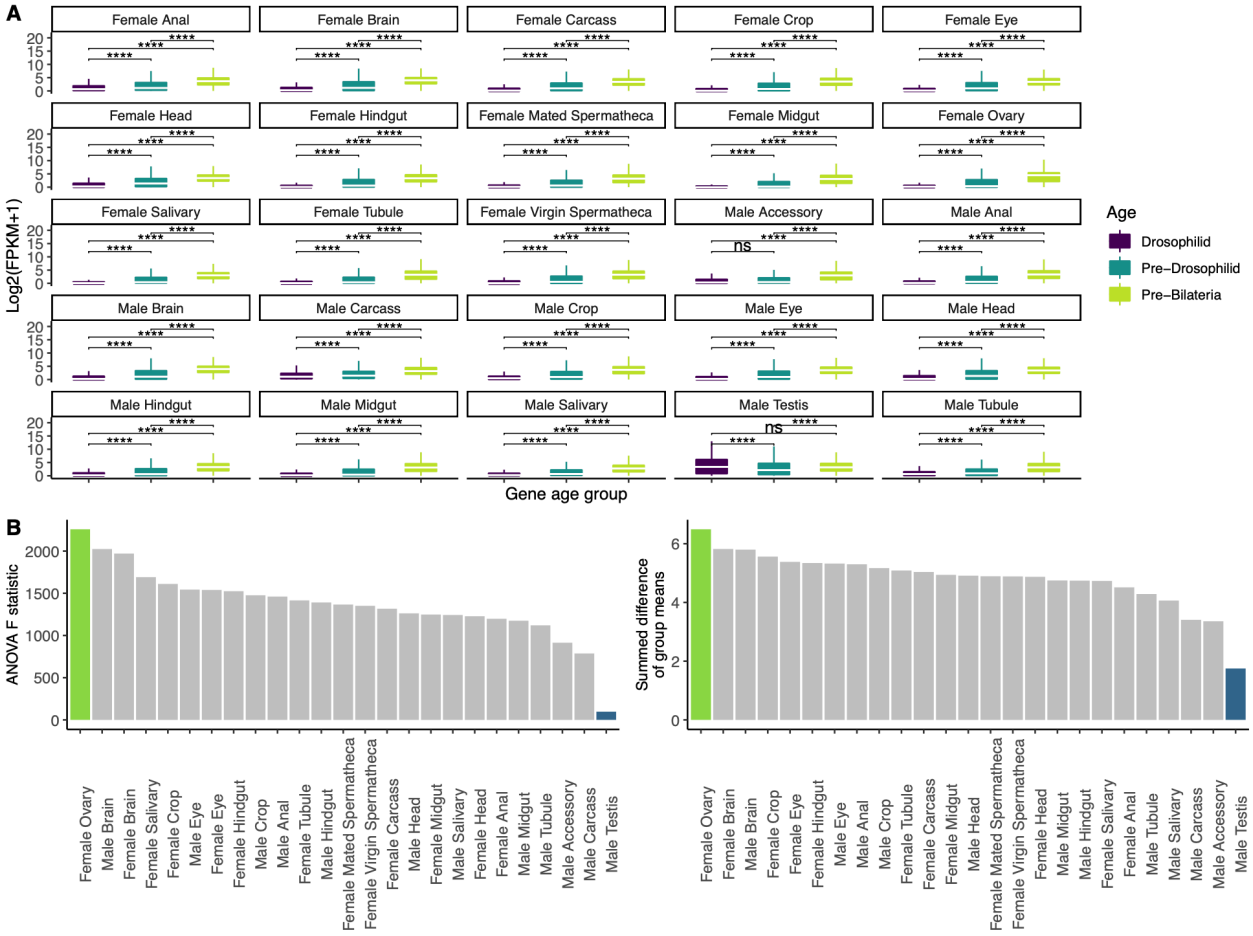
Supplemental Figure 3: Young duplicate genes do not confound results from Figure 1. Recently duplicated genes may have high sequence similarity to their parent copies, causing mapping ambiguities. Shown is the analysis from Figure 1, with genes annotated by Kondo et al. as "*melanogaster*-only" removed. No conclusions from Figure 1 are changed, indicating that *melanogaster*-specific genes do not confound the high testis-specificity of Drosophilid genes.



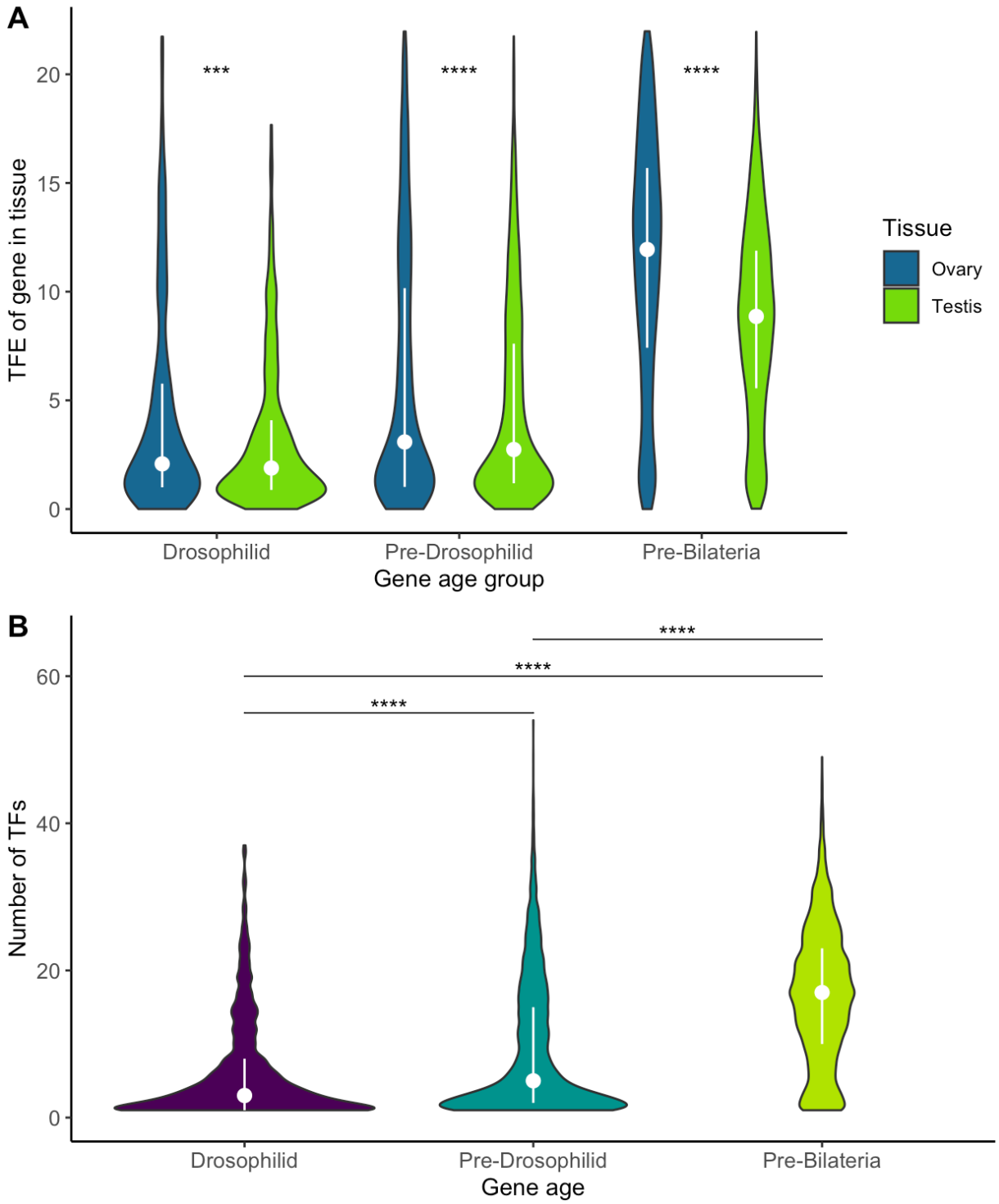
Supplemental Figure 4: Alternate gene age assignments confirm patterns of tissue bias and tissue specificity. The main figures assign Drosophilid genes as those characterized by Zhang et al. and use ages from Kondo et al. for older genes. We remade Figures 1A and 1B with all gene ages assigned from Kondo et al. and found that young genes are more tissue-specific than old genes, young genes are often testis-biased and old genes are ovary biased.



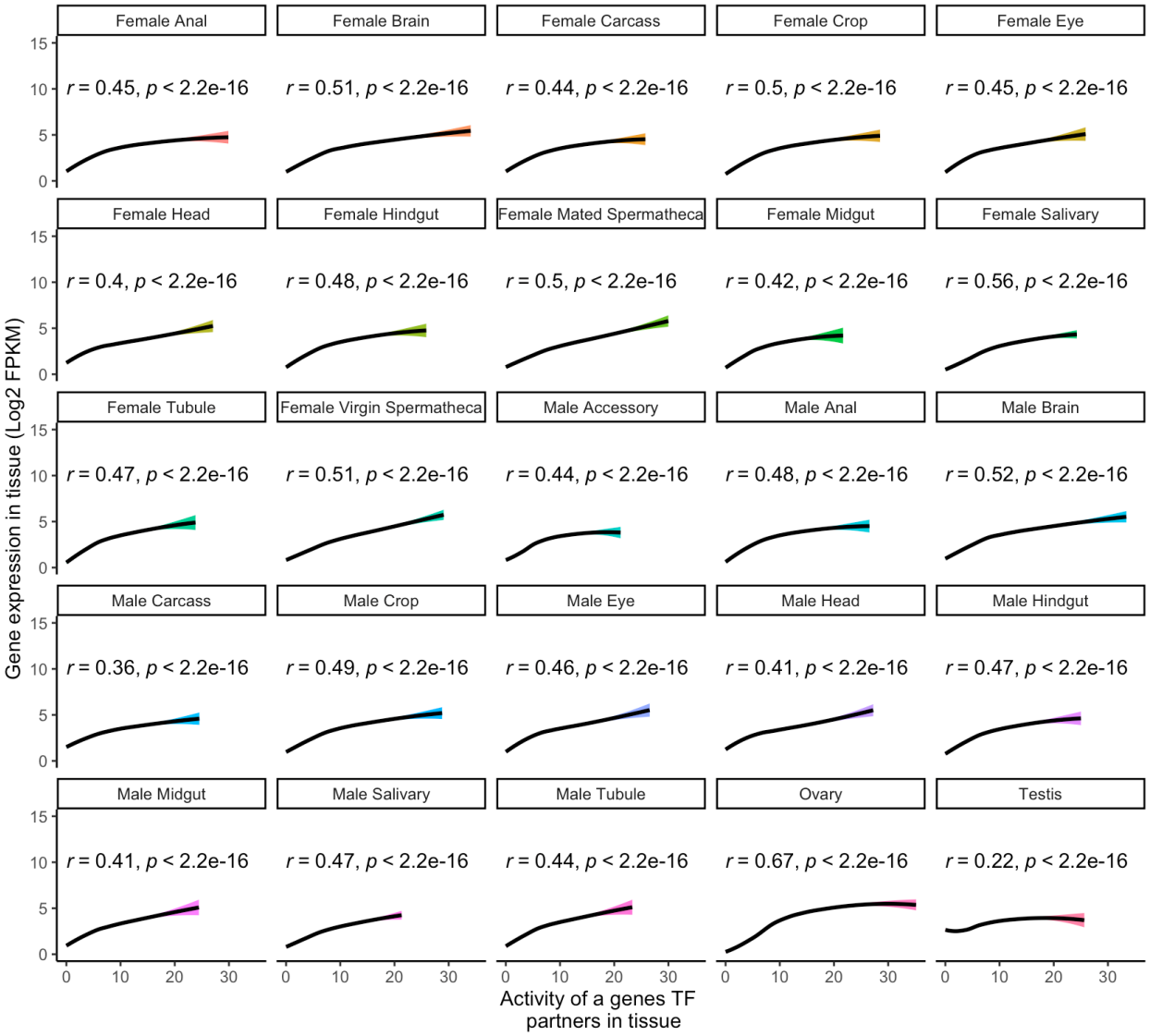
Supplemental Figure 5: Very young genes do not confound results from Figure 2b. With *melanogaster*-specific genes removed, gene expression varies the least between age groups in testis, and most in ovary, just as in the main text. A.) ANOVA F statistic between tissues. B.) Summed difference of mean expression between age groups in each tissue.



Supplemental Figure 6: Alternate gene age assignments do not alter main conclusions from Figure 2. This is the analysis from Figure 2 using only gene age assignments calculated by Kondo et al. In part A, two things have changed. Drosophilid genes have statistically similar expression to Pre-Drosophilid genes in accessory glands, the other male reproductive tissue. In testis, Drosophilid genes are statistically similar to Pre-Bilateria genes instead of Pre-Drosophilid genes. Neither of these changes affect the results from part B, which, like the main figures, shows that ovarian gene expression varies more with age than any other tissue, and testis gene expression varies the least.



Supplemental Figure 7: Alternate gene age assignments do not affect results from Figure 3A and 3B. Using gene ages calculated by Kondo et al., we found that no age group of genes shows elevated TF expression in testis compared to ovary (A). Additionally, the promoters of young genes are less likely to be bound by known TFs than older genes (B).



Supplemental Figure 8: TF expression vs gene expression in every Flyatlas2 tissue. Corresponding to Figure 3, this figure compares TF expression to gene expression across all tissues. Gene expression is least responsive to TF expression in testis, and most responsive to TF expression in ovary. Ovary also has the largest range of TF expression. Lines are a smoothed loess regression with 95 percent confidence intervals. Pearson's r is shown for every tissue, and it is largest in ovary and smallest in testis.

Tissue	Group 1	Group 2	p	p.adj
Female Anal	Drosophilid	Pre-Drosophilid	1.24E-22	9.30E-21
Female Anal	Drosophilid	Pre-Bilateria	1.06E-251	7.95E-250
Female Anal	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00

Female Brain	Drosophilid	Pre-Drosophilid	8.22E-69	6.17E-67
Female Brain	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Brain	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Carcass	Drosophilid	Pre-Drosophilid	3.74E-69	2.81E-67
Female Carcass	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Carcass	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Crop	Drosophilid	Pre-Drosophilid	2.57E-76	1.93E-74
Female Crop	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Crop	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Eye	Drosophilid	Pre-Drosophilid	6.49E-79	4.87E-77
Female Eye	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Eye	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Head	Drosophilid	Pre-Drosophilid	1.04E-69	7.80E-68
Female Head	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Head	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Hindgut	Drosophilid	Pre-Drosophilid	1.55E-67	1.16E-65
Female Hindgut	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Hindgut	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Mated Spermatheca	Drosophilid	Pre-Drosophilid	5.90E-60	4.43E-58
Female Mated Spermatheca	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Mated Spermatheca	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Midgut	Drosophilid	Pre-Drosophilid	7.22E-50	5.42E-48
Female Midgut	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Midgut	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Salivary	Drosophilid	Pre-Drosophilid	4.82E-67	3.62E-65
Female Salivary	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Salivary	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Tubule	Drosophilid	Pre-Drosophilid	1.69E-45	1.27E-43
Female Tubule	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Tubule	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Female Virgin Spermatheca	Drosophilid	Pre-Drosophilid	7.72E-56	5.79E-54
Female Virgin	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00

Spermatheca				
Female Virgin Spermatheca	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Accessory	Drosophilid	Pre-Drosophilid	6.98E-04	5.24E-02
Male Accessory	Drosophilid	Pre-Bilateria	1.22E-156	9.15E-155
Male Accessory	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Anal	Drosophilid	Pre-Drosophilid	1.89E-42	1.42E-40
Male Anal	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Anal	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Brain	Drosophilid	Pre-Drosophilid	1.39E-64	1.04E-62
Male Brain	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Brain	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Carcass	Drosophilid	Pre-Drosophilid	1.83E-12	1.37E-10
Male Carcass	Drosophilid	Pre-Bilateria	3.94E-189	2.96E-187
Male Carcass	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Crop	Drosophilid	Pre-Drosophilid	2.83E-37	2.12E-35
Male Crop	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Crop	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Eye	Drosophilid	Pre-Drosophilid	1.89E-69	1.42E-67
Male Eye	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Eye	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Head	Drosophilid	Pre-Drosophilid	3.62E-67	2.72E-65
Male Head	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Head	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Hindgut	Drosophilid	Pre-Drosophilid	1.37E-28	1.03E-26
Male Hindgut	Drosophilid	Pre-Bilateria	7.48E-298	5.61E-296
Male Hindgut	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Midgut	Drosophilid	Pre-Drosophilid	2.62E-33	1.97E-31
Male Midgut	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Midgut	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Salivary	Drosophilid	Pre-Drosophilid	9.26E-39	6.95E-37
Male Salivary	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Salivary	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Male Tubule	Drosophilid	Pre-Drosophilid	2.87E-14	2.15E-12
Male Tubule	Drosophilid	Pre-Bilateria	1.68E-234	1.26E-232
Male Tubule	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00

Ovary	Drosophilid	Pre-Drosophilid	1.32E-36	9.90E-35
Ovary	Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Ovary	Pre-Drosophilid	Pre-Bilateria	0.00E+00	0.00E+00
Testis	Drosophilid	Pre-Drosophilid	5.47E-13	4.10E-11
Testis	Drosophilid	Pre-Bilateria	7.86E-01	1.00E+00
Testis	Pre-Drosophilid	Pre-Bilateria	7.60E-74	5.70E-72

Supplemental table 1: Raw and adjusted p values for Figure 2. For each tissue and pairwise comparison shown in Figure 2. These are the raw and Bonferroni-corrected p values for each comparison.