Transcriptional and mutational signatures of the aging germline

Evan Witt, Christopher B Langer, Nicolas Svetec, Li Zhao*

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

*Correspondence to: lzhao@rockefeller.edu

Abstract

Aging is a complex biological process that is accompanied by changes in gene expression and mutational load. In many species, including humans, older fathers pass on more paternally-derived *de novo* mutations; however, the cellular basis and cell types driving this pattern are still unclear. To explore the root causes of this phenomenon, we performed single-cell RNA-sequencing (scRNA-seq) on testes from young and old male *Drosophila,* as well as genomic sequencing (DNA-seq) on somatic tissues from the same flies. We found that early germ cells from old and young flies enter spermatogenesis with similar mutational loads, but older flies are less able to remove mutations during spermatogenesis. Mutations in old cells may also increase during spermatogenesis. Our data reveal that old and young flies have distinct mutational biases. Many classes of genes show increased post-meiotic expression in the germlines of older flies. Late spermatogenesis-enriched genes have higher dN/dS than early spermatogenesis-enriched genes, supporting the hypothesis that late spermatogenesis is a source of evolutionary innovation. Surprisingly, young fly enriched genes show higher dN/dS than old fly enriched genes. Our results provide novel insights into the role of the germline in *de novo* mutation.

Introduction

Aging is a process that is accompanied by phenotypic changes in animals. These phenotypic changes include both observable traits and intermediate traits, such as gene expression. Aging can also impact the health of offspring and evolution when it occurs in the reproductive period, such as passing a higher amount of *de novo* mutations to the offspring. Most novel mutations are inherited from the paternal germline, and the number of mutations inherited increases with paternal age (Crow, 2000; Gao et al., 2016, 2019). Some studies have attributed excess paternal mutations to the increased number of cell divisions that cycling spermatogonial

32 stem cells undergo throughout the life of the male (Drost and Lee, 1995; Gao et al., 2011; Li et
33 al., 1996). Conversely, other reports have found that the excess cell divisions do not track the
34 ratio of maternal to paternal mutations during aging (Gao et al., 2019; Huttley et al., 2000),
35 suggesting instead that lifestyle, chemical and environmental factors cause this discrepancy
36 (Irigaray et al., 2007; Parkin et al., 2011). Previous studies of the effect of age on paternally
37 inherited mutations have inferred *de novo* mutations through sequencing of parents and offspring
38 (Gao et al., 2016). These methods are highly useful, but they only capture *de novo* mutations that
39 have evaded repair mechanisms, ended up inside a viable gamete, fertilized an egg, and created a
40 viable embryo. Much less is known about the dynamics of mutation and repair inside the male
41 germline. One study found that mutations arise least frequently in human spermatogonia (Moore
42 et al., 2021). In our previous work (Witt et al., 2019), however, we instead found that mutational
43 load is highest in the earliest stages of spermatogenesis. Taken together, these results imply that
44 most mutations occur prior to germline stem cell (GSC) differentiation and are removed during
45 spermatogenesis. But are these mutations replicative in origin? If so, we would expect germline
46 stem cells from older flies to be more mutated than those from younger flies.
47 In addition to these mutational effects, aging is known to cause other germline
48 phenotypes such as lower numbers of germ cells and reduced germline stem cell proliferative
49 capacity (Lee et al., 2020). The GSC microenvironment also undergoes chemical changes
50 associated with reductions in fecundity (Jones, 2007). These phenotypic consequences, in fact,
51 could be linked to mutation, as the germline mutational rate in young adults correlates with
52 longevity (Cawthon et al., 2020). As such, the germline mutation rate has consequences for both
53 an organism and its descendants.
54 In our previous study (Witt et al., 2019), we used single-cell RNA-sequencing (scRNA-
55 seq) to follow germline mutations throughout *Drosophila* spermatogenesis and found evidence
56 that germline mutations decline in abundance throughout spermatogenesis. We also found
57 evidence that some germline genome maintenance genes are more highly expressed in GSCs and
58 early spermatogonia, the earliest male germ cells. Our results were in line with the idea that
59 active DNA repair plays a role in the male germline (Xia et al., 2020), but since age is an
60 important factor for mutational load, in this study we directly compare patterns in young and old
61 testis.

To study transcriptional and mutational signatures in the aging germline, we generated scRNA-seq data from *Drosophila melanogaster* testes 48 hours and 25 days after eclosion ("Young" and "Old" respectively). We also sequenced correlated genomic DNA from each sample to confirm that each detected mutation was a real *de novo* germline mutation. Our results support our previous observation that the proportion of mutated cells declines throughout spermatogenesis for young flies. For old flies, however, the proportion of mutated cells begins high and remains high throughout spermatogenesis. We found that on a molecular level, each class of older germ cell has a higher mutational burden than comparable cells from young flies. Additionally, older flies carry a higher proportion of C>G and C>A mutations. Our results indicate that the old germline is more highly mutated and transcriptionally dysregulated. We did not find evidence of a profound shift in the expression of genome maintenance genes; however, a number of these genes were highly expressed in young and old flies. We also find that patterns of global gene expression differ between young and old testes, including increased post-meiotic expression of *de novo* genes, TEs, and canonical genes. We found that early spermatogenesis-enriched genes have lower dN/dS than late spermatogenesis-enriched genes, and that genes enriched in older germ cells have lower dN/dS than genes enriched in younger germ cells. These findings provide a deeper insight into the process of spermatogenesis as a key source of *de novo* mutations.

Results

A cell atlas of the aged male germline

We aimed to capture representative cell types from the major somatic and germline cell types of the testis (Figure 1A). We generated testes scRNA-seq data from male flies 48 hours (young) and 25 days (old) after eclosion to facilitate the identification of *de novo* mutations (Figure 1B, Supplemental Table 1). Each of the six libraries was made with approximately 30 pairs of fly testes. We used Cellranger (Zheng et al., 2017) to align these libraries against the FlyBase (Thurmond et al., 2019) *D. melanogaster* genome (version R6.32). We used previously established marker genes (Witt et al., 2021a) to annotate cell types for the young and old flies separately. Dot plots showing the expression of key marker genes are shown in Supplemental Figure 1. To confirm that mutations observed are from the germline, we prepared and sequenced

somatic genomic DNA libraries from the carcasses of the same flies used for scRNA-seq and used these samples as control.

Using Seurat 4 (Satija et al., 2015), we classified somatic cells into four broad types: hub cells, cyst cells, accessory gland, and epithelial cells. We split germ cells into six types, listed from earliest to latest: germline stem cells/early spermatogonia, late spermatogonia, early spermatocytes, late spermatocytes, early spermatids, and late spermatids. In total, we characterized 23489 cells from young flies and 28861 cells from old flies (Figure 1C, Supplemental Table 1). We found that for each age group and cell type, the 3 replicates from each age group largely corroborate each other, with Pearson's r values over 0.91 between replicates and cells of the same age (Supplemental Figure 2). After cell type assignments, we used Seurat 4 to perform downstream analyses on the integrated dataset.

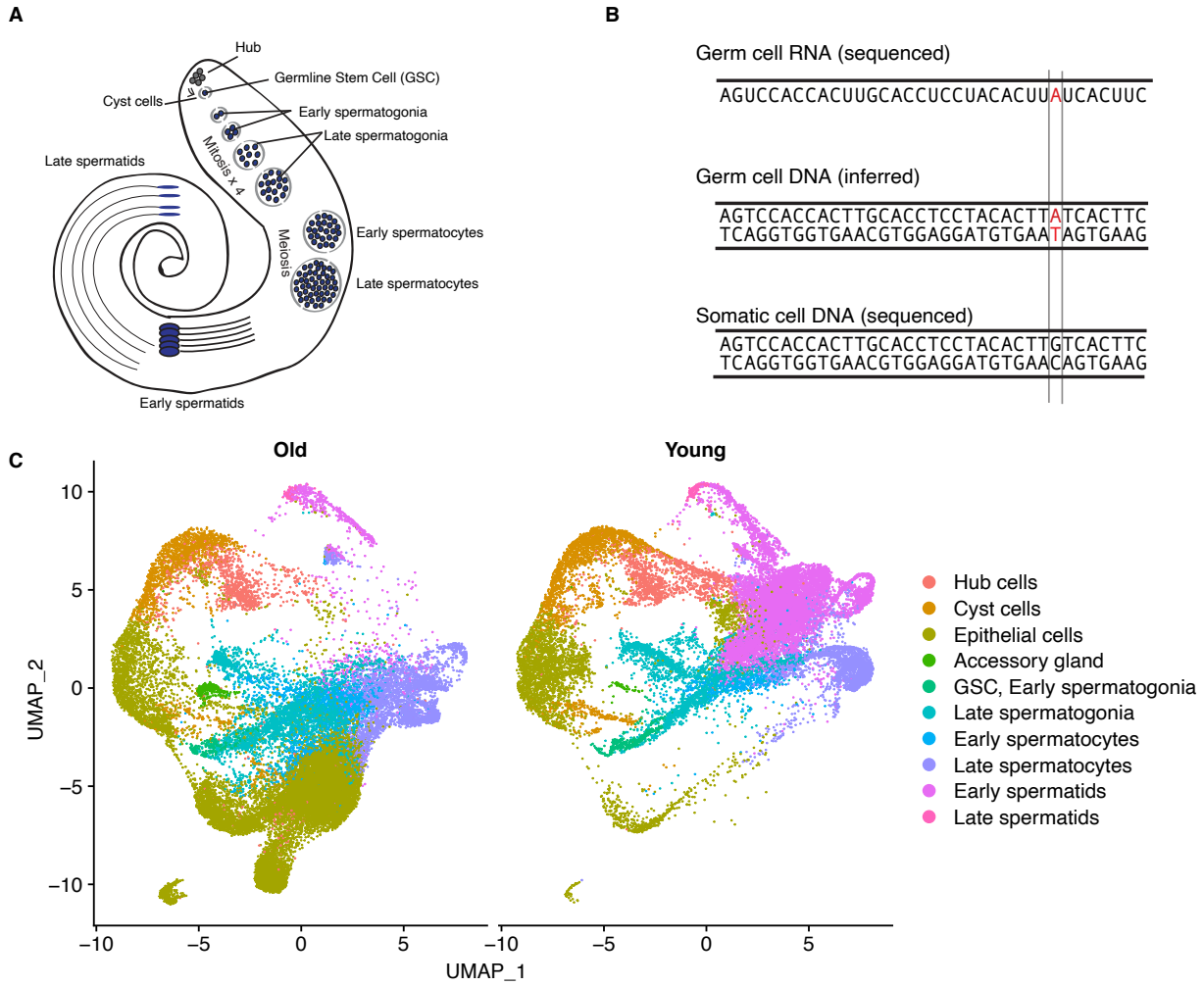


Figure 1: Overview of experimental design and visualization of old and young datasets.

105 A) Diagram of *Drosophila* testis cell types and the marker genes used to identify each cell type. B) experimental
106 rationale: we infer mutated genomic sites in germ cells using scRNA-seq data. If the same locus is unmutated in
107 somatic cell DNA, we call the SNP in red a *de novo* mutation. We can detect a mutation if it is present on both
108 strands or only the template strand. C) dimensional reduction showing the cell-type assignments of scRNA-seq data
109 from young and old flies.
110

111 **Older flies show impaired mutational repair during spermatogenesis**

112 We identified germline SNPs in each sample and matched them to every cell with reads
113 corroborating a given SNP (Figure 2). To assess the mutational burden between young and old
114 flies, we compared, for each cell type, the proportion of cells where at least one mutation was
115 detected. We also compared, for every cell type, the number of detected SNPs per unique
116 molecular index (UMI) to account for differences in coverage between libraries. We did not find
117 any SNPs in more than one replicate, indicating that recurrent age-related transcriptional errors
118 or RNA editing events did not bias our results.

119 In young flies, the proportion of mutated cells declines drastically during
120 spermatogenesis, indicating that lesions are either repaired or that mutated cells are removed
121 from the population. Old flies begin spermatogenesis with a similar proportion of mutated
122 GSC/early spermatogonia, but their mutational burden remains high throughout spermatogenesis.
123 Proportions of mutated cells are statistically similar for young and old flies in GSC/Early
124 spermatogonia but begin to diverge in later cell types. By the end of spermatogenesis, young
125 flies achieve a lower proportion of mutated cells compared to older flies, whose proportion of
126 mutated cells remains high throughout.

127 Old spermatocytes and spermatids have significantly higher proportions of mutated cells
128 (Figure 2A). To confirm that this trend was not confounded by different read depths across
129 replicates, we counted the number of detected SNPs per Unique Molecular Index (UMI) for
130 every cell type and found that RNA molecules from old cells are consistently more likely to
131 carry mutations than young cells of the same type (Figure 2B). This suggests that much of the
132 elevated mutational load of the older germline occurs before spermatogenesis, accumulated
133 within cycling germ cells. However, we also observe that later germ cells from old flies have
134 more mutations per RNA molecule than older GSC/early spermatogonia. This observation would
135 be expected if some germline mutations arose during spermatogenesis.
136

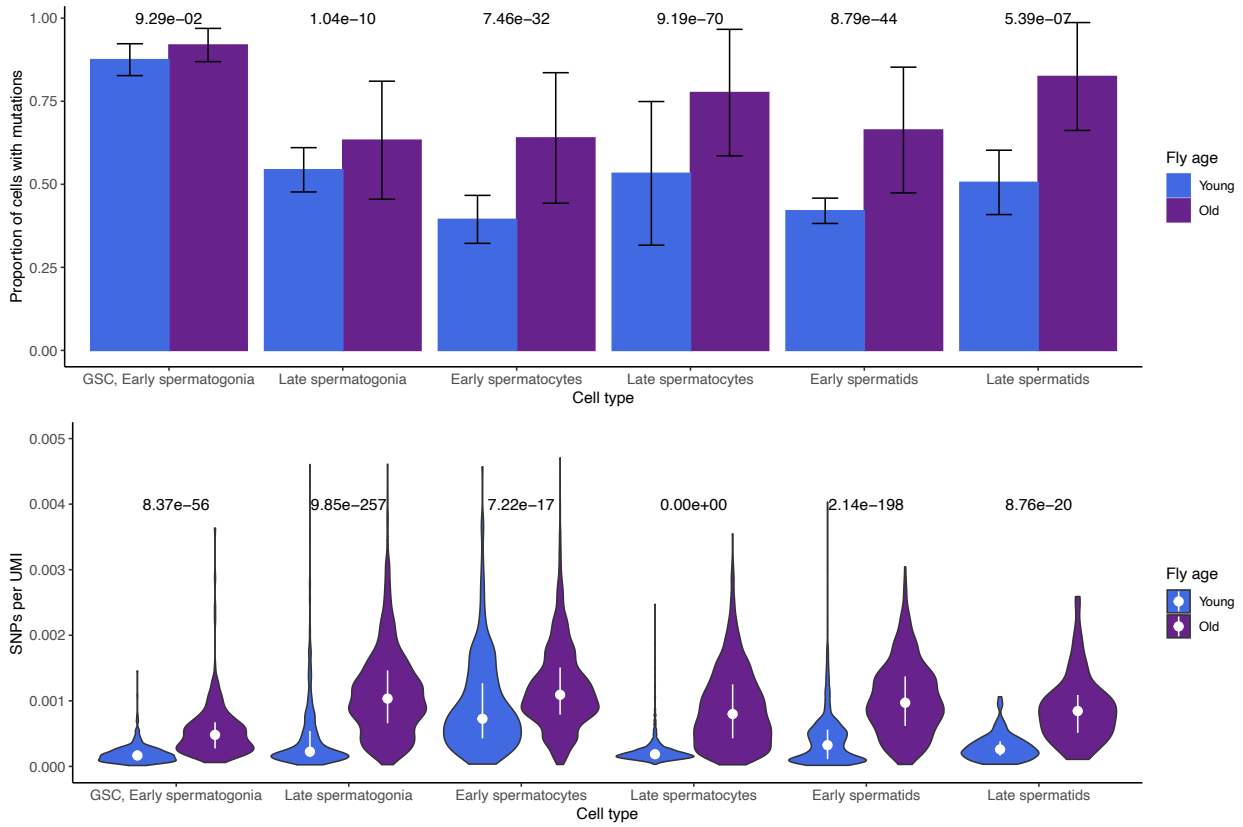
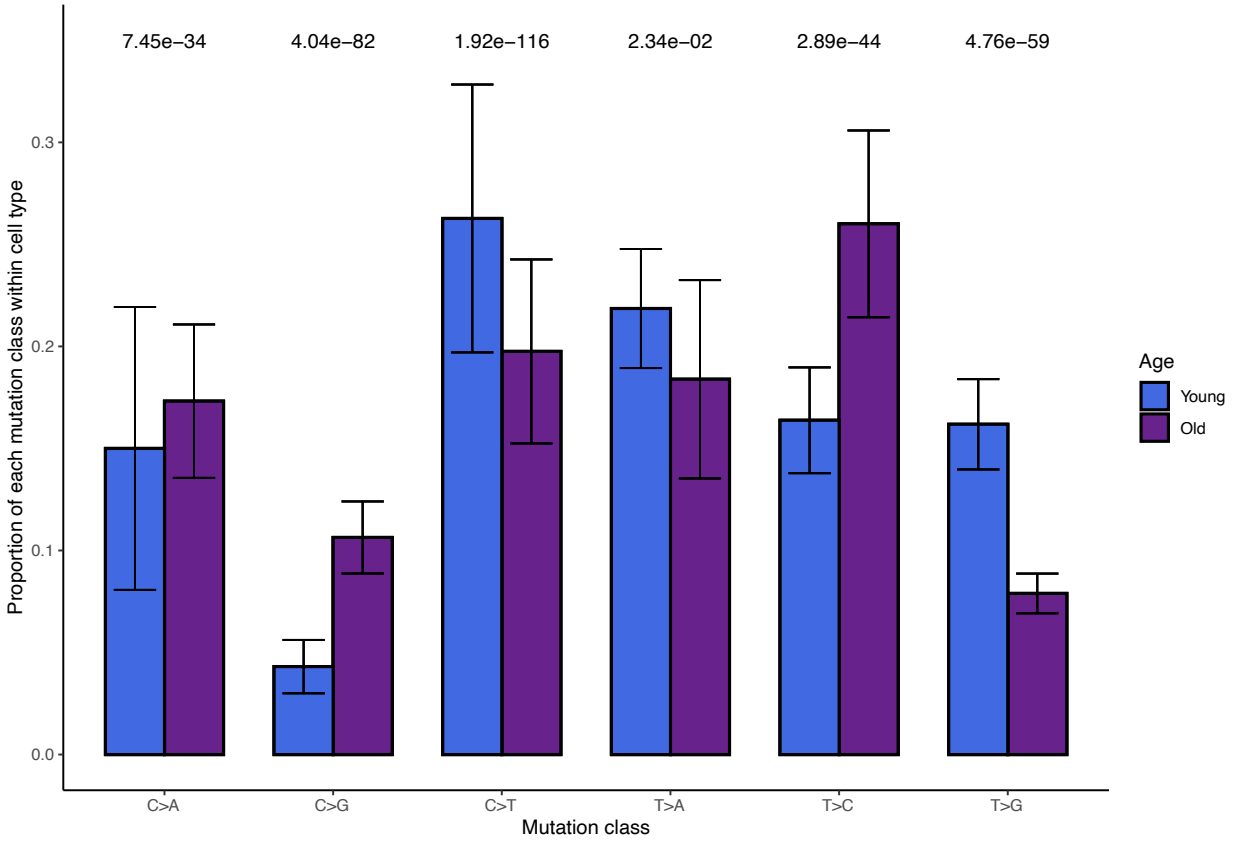


Figure 2: The proportion of mutated cells declines more for young flies than old flies, while old flies have a higher mutation load. A) For old and young flies and every cell type, shown are the proportions of cells of each type carrying at least one mutation. Error bars are standard error. P values are Bonferroni-corrected from a chi-square test of mean proportions between young and old cells of a type. The proportion of mutated cells declines more for young flies than not old flies, during spermatogenesis. B) for each cell, the number of SNPs divided by the number of Unique Molecular Indices (UMIs) detected, a proxy for read depth. P values are Bonferroni-corrected from a chi-square test of proportions comparing young and old cells of a type. Ever class of older germ cells has more mutations per RNA molecule detected. This indicates that the higher mutational load of older flies precedes spermatogenesis, and may increase during spermatogenesis.

T>C and C>G substitutions are enriched in old flies

We compared the relative proportions of the six major classes of mutation between young and old flies. Young flies and old flies have distinct mutational signatures. In young flies, we found that C>T and T>G mutations are enriched compared to old flies (Figure 3). Using a chi-square test of proportions, we found that old flies were significantly enriched for T>C and C>G mutations. This suggests an age-related mutational or repair bias during spermatogenesis. We

155 asked whether these mutational signatures were due to differential activity of genome
156 maintenance genes (Svetec et al., 2016; Witt et al., 2019) and found that, as a group, genome
157 maintenance genes are similarly expressed in most cell types (Supplemental Figure 3). As such,
158 the mutational load of older flies cannot be purely attributed to age-related global
159 downregulation of genomic maintenance genes, although it is likely that the regulation of
160 genome maintenance genes at the protein level differs between young and old testis.
161
162



163
164 **Figure 3: Age-related trends in mutational signatures.** For young and old flies, shown are the relative proportions
165 of the 6 types of mutations (each class is equivalent to a complementary mutation, for example, T>G also represents
166 A>C). Error bars are standard error. Bonferroni-corrected P values are from a chi square test of proportions
167 comparing the mean proportions between young and old flies. T>C and C>G mutations are enriched in old flies,
168 while C>T and T>G mutations are underrepresented.
169

170 **Many genome maintenance genes are more highly expressed in early germ cells from**
171 **young flies**

172 We performed differential expression testing between every cell type (Table 1) and
173 focused on a list of 211 genes related to DNA damage repair compiled from our previous work
174 (Svetec et al., 2016; Witt et al., 2019). We found that in GSC/early spermatogonia, 7 genome
175 maintenance genes were more highly expressed in young flies and 14 more highly expressed in
176 old flies (Supplemental Figure 3, Supplemental Table 2). In spermatocytes and spermatids,
177 comparatively few genes are differentially expressed between young and old testes. This
178 corroborates our earlier observation that genome maintenance genes are generally less expressed
179 in old GSC/early spermatogonia, which is also the most mutated cell type in both old and young
180 flies. Depleted expression of genome maintenance genes in the earliest germ cells could impact
181 the efficiency of germline DNA repair throughout spermatogenesis.
182 Two transcription-related genes, Rbp8 (FBgn0037121) and RpII15 (FBgn0004855) are
183 more highly expressed in young GSC/Early spermatogonia than in old flies. Rbp8, also known as
184 B52, is essential for DNA topoisomerase I recruitment to chromatin during transcription (Juge et
185 al., 2010). DNApol-iota (FBgn0037554), which is a gene involved in translesion synthesis
186 (Ishikawa et al., 2001) and may be important UV damage response (Svetec et al., 2016), was
187 highly expresssed in young GSCs. This suggests that our observed mutational signatures of older
188 flies might be caused by defects in transcription-coupled repair or DNA damage related repair,
189 although the mechanisms of germline genomic surveillance are yet to be fully understoods. In
190 old flies, lower expression of RpII15, also known as RNA Polymerase II, subunit I, might further
191 explain reduced transcription in these cell types from old flies. The gene 14-3-3epsilon
192 (FBgn0020238) is highly expressed in old GSC and spermatogonia, which may play a role in cell
193 division and apoptosis in early germ cells for old testis (Su et al., 2001). The cellular-level
194 mutational signature (Supplemental Figure 4) is in line with previous work suggesting that
195 highly expressed genes evolve more slowly (Drummond et al., 2005; Good and Nachman, 2005).
196

Cell type	# genes enriched in young flies	# genes enriched in old flies
GSC, Early spermatogonia	273	251
Late spermatogonia	144	337
Early spermatocytes	40	367
Late spermatocytes	80	281
Early spermatids	167	277
Late spermatids	52	68

197 **Table 1: Numbers of age-enriched genes per cell type.** Arranged from top (early) to bottom (latest). For young
198 flies, the cell type with the most enriched genes is GSC/Early spermatogonia, the earliest germ cell class. Late

199 spermatids are the most mutated class of cell in older flies. In every class of cells except GSC, early and late
200 spermatogonia, old flies have more enriched genes than young flies.
201
202 ***De novo* genes, transposable elements, and other genes show increased post-meiotic**
203 **expression in older flies.**
204 In our previous work we found that *de novo* genes are highly enriched in meiotic cells
205 (Witt et al., 2019). We asked whether transcriptional dysregulation of the aging germline could
206 impact the expression of genetic novelties such as *de novo* genes and transposable elements
207 (TEs). We performed a parallel analysis of our scRNA-seq data with a custom reference
208 containing 267 testis-expressed *de novo* genes identified from our 2019 study, as well as 239 TEs
209 (Lawlor et al., 2021). We then scaled expression of every gene, centered at zero to compare the
210 expression patterns of groups of genes (Figure 4).
211 Overall, the expression patterns for *de novo* genes resemble those of other genes, and
212 both *de novo* genes and other genes are more expressed in the late spermatids of older flies
213 (Figure 4A, $p = 7.6e-10$, $p < 2e-16$, Supplemental Table 3). This suggests that *de novo* genes
214 show similar expression regulation related to aging compared to old genes and that the regulatory
215 environment acts similarly to conserved and young genes.
216 We were also interested in the expression of TEs, since TE suppression is important for
217 germ cell development (Lee and Langley, 2010). In other tissues, such as the aging brain, the
218 amount of certain types of TEs change with age (Li et al., 2013). We found TE expression is also
219 globally enriched in older late spermatids (Figure 4B, Supplemental Table 4), however this
220 pattern is not more extreme compared to annotated genes. Since gene expression is supposed to
221 stop after meiosis, it is possible that elevated gene expression after meiosis is a consequence of
222 reduced post-meiotic transcriptional suppression. Although all 3 classes of genes were also more
223 highly expressed in the late spermatogonia of young flies, this increase was strikingly large for
224 transposable elements ($p = 0.0018$). This result corroborates earlier work which found a similar
225 burst of transposon activity in early spermatogenesis, likely when spermatogonia transit to
226 spermatocytes (Lawlor et al., 2021). As such, the reduced early TE expression in older germlines
227 may reflect a dysregulated transcriptional environment.
228

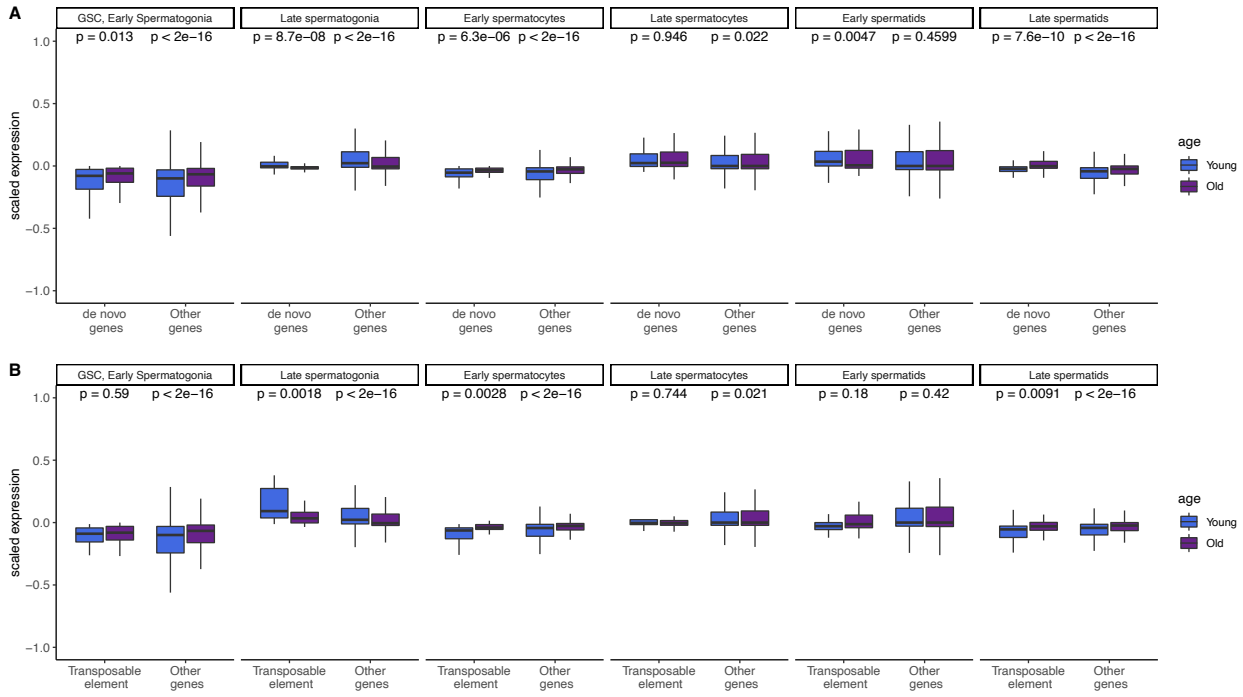


Figure 4: Global expression patterns of *de novo* genes and transposable elements changes with age in each cell type. A.) Scaled expression of *de novo* genes and other genes (not including TEs) across cell types. Expression of both gene types is enriched in the late spermatids of older flies. B.) Scaled expression of TEs and other genes (not including *de novo* genes) across cell types. Transposable element expression is highly enriched in late spermatogonia of young flies, and enriched in the late spermatids of older flies. Figure shows p values by Wilcoxon rank sum tests.

Early spermatogenesis-biased genes have lower dN/dS than late spermatogenesis-biased genes

We asked whether functional constraint varies for age-specific or cell-type specific genes. We defined "age-biased" genes as genes differentially expressed between old and young flies in the same cell type. We defined "stage-biased" genes as genes differentially expressed between cell types of the same age. To find stage-biased genes, we split our dataset into "old" and "young" cells and then performed Seurat's FindMarkers function between GSC/early spermatogonia (early germline) and late spermatids (late germline) for each age group. Using dN/dS data from flyDIVaS (Stanley and Kulathinal, 2016), we compared dN/dS values for early-stage-biased and late-stage-biased genes. In both young and old flies, the genes with expression bias toward later stages exhibit higher dN/dS than genes biased in GSC/early spermatogonia (Figure 5A), which is in line with the idea that genes expressed in late spermatogenesis may

evolve rapidly. Note that spermatocytes and spermatids are also hotspots for the expression of novel genes including *de novo* originated genes. Our findings are also similar to a murine study which found that genes expressed in early spermatogenesis are under more evolutionary constraint than genes expressed late germ cells (Schumacher and Herlyn, 2018).

To identify age-biased genes, we divided the dataset by cell type and identified genes enriched in old or young cells of the same type. We found that in both GSC/early spermatogonia and late spermatids, genes biased in young cells have higher dN/dS than genes biased in old cells (Figure 5B). This result show that many early-germ-cell-biased genes in young flies evolve rapidly, that is, a higher proportion of rapidly evolving genes are enriched in young-age germline stem cells and spermatogonia than in old ones. This suggests that, if antagonistic pleiotropy in aging plays an important role in the shift of gene expression (Austad and Hoffman, 2018; Williams, 1957), rapidly evolving genes are often expressed and function in young animals and are subject to antagonistic pleiotropy.

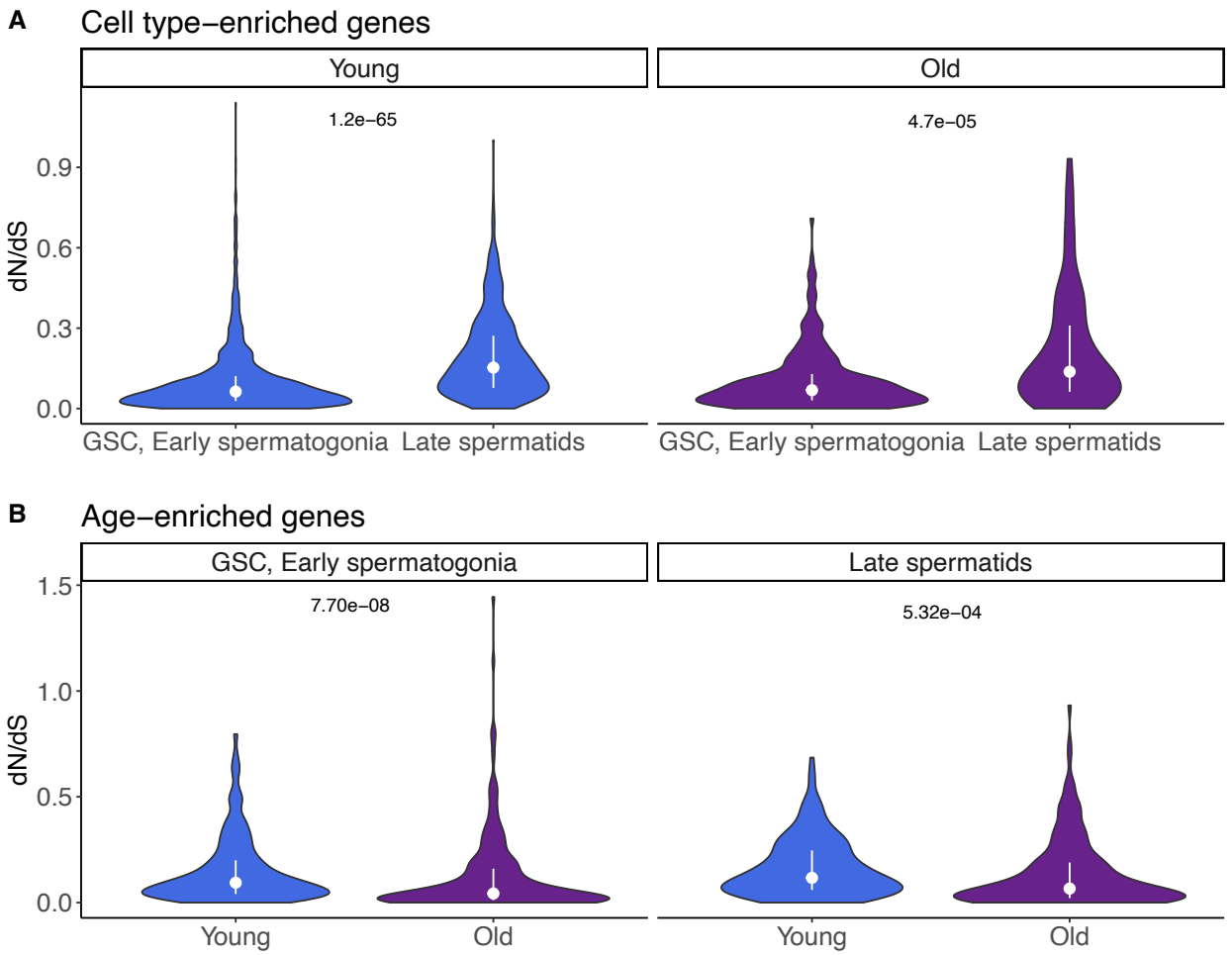


Figure 5: dN/dS trends of cell type-enriched (stage-biased) and age-enriched (age-biased) genes. We calculated gene enrichments in two different ways. First, we calculated gene enrichment between GSC/early spermatogonia and late spermatids separately for young and old flies. Then, we identified genes enriched in young and old cells within a cell type. A) Cell type-enriched genes: in both old and young flies, genes enriched in late spermatids have higher dN/dS than genes enriched in GSC/early spermatogonia (Wilcoxon rank sum test, p values 1.2e-65, 4.7e-5, respectively). B) Age-enriched genes: In GSC/early spermatogonia, but not late spermatids, young-enriched genes have higher dN/dS than old-enriched genes (Wilcoxon rank sum test, p values 7.70e-08 and 5.32e-04, respectively). dN/dS values are from flyDIVaS.

Discussion

Mutational load is an equilibrium between mutation and repair. In old flies, this equilibrium may shift away from repair. In this work, we show that the germline of older flies is less able to remove *de novo* mutations compared to the germline of young flies. Throughout spermatogenesis, we observed that germ cells from older flies have more mutations per RNA

278 molecule than comparable cells from younger flies. This finding adds a new explanation for the
279 still-controversial mechanism behind the increased age-dependent mutational load of the male
280 germline. Our finding of increased mutational load in older GSC/early spermatogonia suggests
281 that much of the mutational load of older flies accumulates prior to spermatogenesis. However,
282 later germ cells of older flies have more SNPs per UMI, suggesting that mutations may
283 accumulate during spermatogenesis. Our work corroborates previous work that found that the
284 huge excess of male germline divisions is too large to explain the much smaller ratio of
285 male/female-inherited mutations during parental aging (Gao et al., 2019). Our finding that early
286 germ cells from young and old flies are similarly mutated supports the notion that many age-
287 related germline mutations are not due to replicative processes. Some of our conclusions are in
288 line with recent findings that germline stem cells have the lowest mutation rate of any human cell
289 type (Moore et al., 2021).

290 In addition to being highly mutated, the older germline shows distinct mutational
291 signatures compared to the younger germline. For example, we found a statistical
292 overrepresentation of C>A and C>G mutations in old flies and an underrepresentation of C>T
293 mutations. These altered ratios of single nucleotide polymorphisms could be caused by
294 differential activity of DNA repair pathways in the old germline. We did not find strong evidence
295 of global downregulation of genome maintenance gene expression but found lower expression of
296 a few key transcriptional genes in GSC of older flies. Altered types and numbers of *de novo*
297 mutations would likely have implications for a population, affecting the type and frequency of
298 genetic novelties that emerge (Loewe and Hill, 2010). In the future, it would be interesting to
299 understand the molecular mechanisms contributing to mutational bias in germ cells.
300 Additionally, these methods should be reproduced with mated flies to examine if germline
301 mutational bias can be affected by a male's reproductive activity.

302 We observed that scaled expression of all genes is generally down in GSC/early
303 spermatogonia but up in late spermatids. The latter result is intriguing because transcription
304 largely ceases after meiosis in the male *Drosophila* germline (Barreau et al., 2008). While the
305 downregulation of transcription in early germ cells could have implications for germline DNA
306 repair, the potential effects of increased post-meiotic transcription are less clear. It could have no
307 effect, or it could affect spermatid maturation or sperm competition, potentially affecting

308 fertility. Indeed, increased male age associates with reduced fertility in humans (Harris et al.,
309 2011).

310 Other studies have proposed that the testis uses ubiquitous gene expression to detect
311 genomic lesions and repair them with transcription-coupled repair (Xia and Yanai, 2022; Xia et
312 al., 2020). Due to our bias towards detecting mutations in expressed genes, our data is not ideal
313 to test this hypothesis. We noted, however, that genes with many detectable SNPs tend to be
314 lowly expressed across replicates (Supplemental Figure 4), a finding that appears to be consistent
315 with the transcriptional scanning model. Broadly, this is in line with previous work that highly
316 expressed genes show lower dN/dS, which previous work show that this was driven by various
317 types of selective pressures (Drummond et al., 2005; Good and Nachman, 2005). Our results
318 surprisingly found this pattern in mutational bias, even before natural selection acts on the genes.
319 This result also supports the notion that it is very unlikely that the germline mutations we
320 observed are directly from transcription errors, as that would lead to more mutations to highly
321 expressed genes. Impaired mutational repair of the older male germline may be partially due to
322 deficiencies in transcription-coupled repair (Deger et al., 2019).

323 The global post-meiotic upregulation of transcription extends beyond conserved genes.
324 We observed that many transposable elements are highly enriched in the early germline of young
325 flies. Not only could this have implications for transposable element mobilization, it could also
326 create heritable changes in chromatin structure, signaling, or gene expression (Chuong et al.,
327 2017; Lanciano and Cristofari, 2020).

328 The global deregulation of gene expression during aging also has interesting implications
329 for evolution. Consistent with prior work (Schumacher and Herlyn, 2018), we found that genes
330 enriched in late germ cells have higher dN/dS than genes enriched in early germ cells for both
331 young and old flies. This result suggests that spermatocytes and spermatids are sources of rapid
332 evolution or positive selection. Considering that spermatocytes and spermatids are also the stages
333 where *de novo* gene are the most abundant and where they likely function (Witt et al., 2019) our
334 results highlight the importance of late spermatogenesis in transcriptional and functional
335 innovation.

336 Unexpectedly, we also found that, within analogous cell types, genes enriched in cells
337 from young flies have higher dN/dS than genes enriched in old flies. There are two possible
338 explanations. First, rapidly evolving genes or genes with more adaptive changes may provide a

339 greater evolutionary advantage to young flies than old flies. Second, since we used the scaled
340 expression, old-enriched genes may reflect a specific set of genes that are essential for aging
341 animals, which may be more likely to be house-keeping genes. Our results are interesting in the
342 light of recent work which found that genes expressed later in life tend to fix nonsynonymous
343 mutations more frequently (Cheng and Kirkpatrick, 2021). One should note that their
344 methodology is different: their age-biased genes were identified from whole-body data, whereas
345 ours were calculated just from male germ cells. Gene expression in the testis is often an outlier
346 compared to other tissues (Witt et al., 2021b), so the results of these two studies are not
347 necessarily in conflict. Nevertheless, the consistent pattern between this study and that of Cheng
348 and Kirkpatrick (Cheng and Kirkpatrick, 2021) is that that genes enriched in late spermatids have
349 a higher dN/dS than those enriched in early germ cells. In this way, at both whole-organism
350 development and germline development level, late-stage biased genes tend to evolve more
351 rapidly.

352 Our study design limits the detection of mutations in expressed transcripts. While we
353 have strict criteria for the identification of novel SNPs, the abundance of false negatives could
354 vary between cell types due to cell-specific variation in transcriptional activity. We are reassured
355 because the most commonly mutated cell type in our datasets is GSC/early spermatogonia,
356 consistent with our previous observations using different datasets and slightly different analytical
357 pipelines. If transcriptional activity biased our inference of mutational load, we would expect
358 spermatocytes, the most transcriptionally active cell type, to appear the most mutated instead.
359 This potential confounder would be resolved if a method became available to simultaneously
360 perform RNA-seq and whole-genome sequencing on thousands of single cells. Current methods
361 may not capture enough cells to comprehensively profile rarer cell types, but we expect this
362 technical challenge to improve with future technological advances.

363 Another potential confounder is that aging might create the appearance of germline SNPs
364 through reduced transcriptional fidelity (Verheijen and van Leeuwen, 2017). We do not think
365 this is a significant source of error, since our SNPs are verified by multiple independent reads
366 and are not allowed to be present in more than one dataset, in fact, we did not observe mutations
367 occur in multiple datasets after filtering, suggesting that hidden common transcription errors or
368 RNA editing do not impact our work. To understand the role that age-related transcriptional
369 fidelity significantly plays in our results, this topic would benefit from high throughput combined

370 scRNA/DNA sequencing from the same cells. The technology allowing us to trace *de novo*
371 mutations throughout the germline is still very new, and we look forward to technological
372 advancements in this exciting field.
373
374 **Methods**
375 **ScRNA-seq library preparation and sequencing**
376 In this study, we used young and old RAL517 for experiments. Briefly, flies were reared
377 in the standard corn syrup medium at room temperature with synchronized 12:12 hour light:dark
378 cycle. Both age groups of flies were kept as virgins after eclosion. Young and old samples were
379 collected 48 hours and 25 days after eclosion, respectively. In detail, virgin males were collected
380 within one hour after eclosion and were transferred to new vials. After 48 hours, we dissected 30
381 pairs of young testes for single-cell suspension and scRNA-seq and kept the carcasses for DNA
382 sequencing. After 25 days, we dissected 30 pairs of old testes for single-cell suspension and
383 scRNA-seq and kept the carcasses for DNA sequencing. We generated three biological replicates
384 for young and old flies, respectively. Flies were all dissected in the morning (ZT1-ZT3 in our lab
385 environment) to reduce expression fluctuation due to circadian rhythm. Single-cell testis
386 suspensions were prepared as described in our previous work (Witt et al., 2019). Libraries were
387 prepared with 10X Chromium 3' V3 kit and sequenced with Illumina Hiseq 4000.
388
389 **Genomic DNA preparation**
390 Fly carcasses were frozen at -80°C, then ground in 200 µL 100 mM Tris-HCl, pH 7.5,
391 100 mM EDTA, 100mM NACl, 0.5% SDS. The mixtures were incubated at 65°C for 40
392 minutes. Then, 160 µl KAc and 240 µl 6M LiCl were added, tubes were inverted 10 times and
393 placed on ice for 30 minutes. Samples were then centrifuged at 18000g at 4°C for 15 minutes.
394 The supernatant was transferred into a new tube and an equivalent volume of isopropanol was
395 added and mixed by inversion. Samples were spun for 15 minutes at 18000g, and the supernatant
396 was discarded. Pellets were washed with 800 µL 70% ethanol and samples were spun at 18000g
397 for 5 minutes, and supernatant discarded. Pellet was air dried for 5 minutes and resuspended in
398 100 µL nuclease-free water. DNA was then sent for Illumina library preparation and sequencing
399 by Novogene.
400

401 **ScRNA-seq data processing**

402 ScRNA-seq data were aligned with Cellranger Count and further processed with Seurat.
403 To assign cell types with Seurat, we used marker genes described in Witt et al. 2021 (Witt et al.,
404 2021a). We normalized the replicates with SCTransform in Seurat and integrated them into a
405 combined Seurat object. We performed clustering and annotation together on all replicates, using
406 normalized counts from the "SC T" slot. Cells were annotated based on the expression patterns
407 of marker genes. Clusters enriched in *bam* and *aub* are GSC/early spermatogonia(Kawase, 2004;
408 Witt et al., 2019, 2021a), and adjacent clusters with less *bam/aub* and less *His2Av* are late
409 spermatogonia. Clusters enriched in *fzo* and *twe* are early and late spermatocytes, respectively
410 (Courtot et al., 1992; Hwa et al., 2002). Clusters with enriched *soti*, but not p-cup are early
411 spermatids (Barreau et al., 2008) and clusters enriched in *p-cup* are late spermatids (Barreau et
412 al., 2008). Clusters enriched in *MtnA* are somatic cells, *dlg1* defines cyst cells (Papagiannouli
413 and Mechler, 2009) and *Fas3* defines hub cells. Epithelial cells are enriched in *MtnA* but not
414 *Rab11* or *Fas3* (Witt et al., 2021a).
415
416 SNPs were called with bcftools (Narasimhan et al., 2016) separately for each single-cell library
417 and each gDNA library. Per-base coverage was calculated for every gDNA sample with
418 Samtools (Li et al., 2009). For each young and old scRNA-seq library, bcftools isec was used to
419 extract mutations only present in the SC data and not the somatic gDNA. Using Samtools, we
420 identified every cell barcode in the scRNA-seq data that corroborated every SNP (details in
421 accompanying code). For each mutated position, we then verified that the corresponding locus in
422 the gDNA file had at least 10 reads supporting the reference allele, and 0 reads supporting the
423 alternative allele. We also required that every SNP be present only in a single scRNA-seq
424 dataset, to reduce the chance that RNA editing events or transcription errors caused us to infer a
425 SNP incorrectly. We also required every SNP to have >=2 reads corroborating it, reducing the
426 potential impact of sequencing errors.
427

428 **Comparisons using scaled expression**

429 To compare gene expression for groups of genes across replicates, we scaled expression
430 using the ScaleData Seurat function separately on each replicate. Expression is scaled such that 0
431 represents a gene's median expression across all cells, 1 represents 1 standard deviation above

that gene's mean expression, and -1 represents 1 standard deviation below. Within a cell type, each gene's scaled expression was averaged between cells. Groups of genes were compared using a two-sample Wilcoxon rank sum test, and p values were adjusted with Bonferroni's correction.

Differential expression testing

For each germ cell type, we made a subset Seurat object containing just that cell type with old and young flies, assigning "age" as the cell identifier. We then used Seurat's FindMarkers function with ident.1 as "Young" and ident.2 as "Old". We classified genes with a Bonferroni-adjusted p value < 0.05 and Log2 fold change > 0 as enriched in young, and <0 as enriched in old. We then constructed volcano plots with the EnhancedVolcano package, while including differentially expressed genes from our list of 211 genome maintenance genes from our previous paper (Witt et al., 2019).

***De novo* gene and TE analysis**

De novo genes from our previous paper (Witt et al., 2019) were added to a reference GTF containing transposable elements from another study (Lawlor et al., 2021). This alternate reference was used to align reads from all libraries with Cellranger. Cell-type annotations were copied from the annotations made for the main Seurat object. Enriched *de novo* genes and TEs were detected with the FindMarkers function in Seurat.

Data availability

Code used for processing of data is deposited at https://github.com/LiZhaoLab/Mutation_project. This repository will include permanent links to large data files including a Seurat RDS and mutation database. Raw sequence data will be released on SRA upon publication.

Funding

The work was supported by NIH MIRA R35GM133780, the Robertson Foundation, a Monique Weill-Caulier Career Scientist Award, a Rita Allen Foundation Scholar Program, and a Vallee

462 Scholar Program (VS-2020-35), and an Alfred P. Sloan Research Fellowship (FG-2018-10627)
463 to L. Z.
464
465 **Acknowledgements**
466 We thank Hong Duan and Connie Zhao at Genomics Resource Center of Rockefeller University
467 for their help with the scRNA-seq libraries, and members of Zhao lab for their helpful comments
468 and suggestions.
469
470 **Declaration of interests**
471 The authors declare no competing interests.
472
473 **Reference**
474 Austad, S.N., and Hoffman, J.M. (2018). Is antagonistic pleiotropy ubiquitous in aging biology?
475 Evol. Med. Public Heal. *2018*, 287–294.
476 Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F., and White-Cooper, H. (2008). Post-
477 meiotic transcription in *Drosophila* testes. Development *135*, 1897–1902.
478 Cawthon, R.M., Meeks, H.D., Sasani, T.A., Smith, K.R., Kerber, R.A., O'Brien, E., Baird, L.,
479 Dixon, M.M., Peiffer, A.P., Leppert, M.F., et al. (2020). Germline mutation rates in young
480 adults predict longevity and reproductive lifespan. Sci. Rep. *10*, 10001.
481 Cheng, C., and Kirkpatrick, M. (2021). Molecular evolution and the decline of purifying
482 selection with age. Nat. Commun. *12*, 2657.
483 Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable
484 elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.
485 Courtot, C., Fankhauser, C., Simanis, V., and Lehner, C.F. (1992). The *Drosophila cdc25*
486 homolog *twine* is required for meiosis. Development *116*, 405–416.
487 Crow, J.F. (2000). The origins, patterns and implications of human spontaneous mutation. Nat.
488 Rev. Genet. *1*, 40–47.
489 Deger, N., Yang, Y., Lindsey-Boltz, L.A., Sancar, A., and Selby, C.P. (2019). Drosophila, which
490 lacks canonical transcription-coupled repair proteins, performs transcription-coupled repair. J.
491 Biol. Chem. *294*, 18092–18098.
492 Drost, J.B., and Lee, W.R. (1995). Biological basis of germline mutation: Comparisons of

493 spontaneous germline mutation rates among drosophila, mouse, and human. Environ. Mol.
494 Mutagen. *25*, 48–64.
495 Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., and Arnold, F.H. (2005). Why highly
496 expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U. S. A. *102*, 14338–14343.
497 Gao, J.-J., Pan, X.-R., Hu, J., Ma, L., Wu, J.-M., Shao, Y.-L., Barton, S.A., Woodruff, R.C.,
498 Zhang, Y.-P., and Fu, Y.-X. (2011). Highly variable recessive lethal or nearly lethal mutation
499 rates during germ-line development of male Drosophila melanogaster. Proc. Natl. Acad. Sci. U.
500 S. A. *108*, 15914–15919.
501 Gao, Z., Wyman, M.J., Sella, G., and Przeworski, M. (2016). Interpreting the dependence of
502 mutation rates on age and time. PLoS Biol. *14*, 1–16.
503 Gao, Z., Moorjani, P., Sasani, T.A., Pedersen, B.S., Quinlan, A.R., Jorde, L.B., Amster, G., and
504 Przeworski, M. (2019). Overlooked roles of DNA damage and maternal age in generating human
505 germline mutations. Proc. Natl. Acad. Sci. *116*, 9491 LP – 9500.
506 Good, J.M., and Nachman, M.W. (2005). Rates of protein evolution are positively correlated
507 with developmental timing of expression during mouse spermatogenesis. Mol. Biol. Evol. *22*,
508 1044–1052.
509 Harris, I.D., Fronczak, C., Roth, L., and Meacham, R.B. (2011). Fertility and the aging male.
510 Rev. Urol. *13*, e184–e190.
511 Huttley, G.A., Jakobsen, I.B., Wilson, S.R., and Easteal, S. (2000). How important is DNA
512 replication for mutagenesis? Mol. Biol. Evol. *17*, 929–937.
513 Hwa, J.J., Hiller, M.A., Fuller, M.T., and Santel, A. (2002). Differential expression of the
514 Drosophila mitofusin genes fuzzy onions (fzo) and dmfn. Mech. Dev. *116*, 213–216.
515 Irigaray, P., Newby, J.A., Clapp, R., Hardell, L., Howard, V., Montagnier, L., Epstein, S., and
516 Belpomme, D. (2007). Lifestyle-related factors and environmental agents causing cancer: An
517 overview. Biomed. Pharmacother. *61*, 640–658.
518 Ishikawa, T., Uematsu, N., Mizukoshi, T., Iwai, S., Iwasaki, H., Masutani, C., Hanaoka, F.,
519 Ueda, R., Ohmori, H., and Todo, T. (2001). Mutagenic and Nonmutagenic Bypass of DNA
520 Lesions byDrosophila DNA Polymerases dpolη and dpolι. J. Biol. Chem. *276*, 15155–15163.
521 Jones, D.L. (2007). Aging and the germ line: where mortality and immortality meet. Stem Cell
522 Rev. *3*, 192–200.
523 Juge, F., Fernando, C., Fic, W., and Tazi, J. (2010). The SR Protein B52/SRp55 Is Required for

524 DNA Topoisomerase I Recruitment to Chromatin, mRNA Release and Transcription Shutdown.
525 PLOS Genet. *6*, e1001124.
526 Kawase, E. (2004). Gbb/Bmp signaling is essential for maintaining germline stem cells and for
527 repressing bam transcription in the *Drosophila* testis. Development *131*, 1365–1375.
528 Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element
529 expression. Nat. Rev. Genet. *21*, 721–736.
530 Lawlor, M.A., Cao, W., and Ellison, C.E. (2021). A transposon expression burst accompanies the
531 activation of Y-chromosome fertility genes during Drosophila spermatogenesis. Nat. Commun.
532 *12*, 6854.
533 Lee, Y.C.G., and Langley, C.H. (2010). Transposable elements in natural populations of
534 Drosophila melanogaster. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1219–1228.
535 Lee, M.-H., Luo, H.-R., Bae, S.H., and San-Miguel, A. (2020). Genetic and Chemical Effects on
536 Somatic and Germline Aging. Oxid. Med. Cell. Longev. *2020*, 4684890.
537 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
538 Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence
539 Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
540 Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., and Dubnau, J.
541 (2013). Activation of transposable elements during aging and neuronal decline in Drosophila.
542 Nat. Neurosci. *16*, 529–531.
543 Li, W.H., Ellsworth, D.L., Krushkal, J., Chang, B.H., and Hewett-Emmett, D. (1996). Rates of
544 nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol.
545 Phylogenet. Evol. *5*, 182–187.
546 Loewe, L., and Hill, W.G. (2010). The population genetics of mutations: good, bad and
547 indifferent. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1153–1167.
548 Moore, L., Cagan, A., Coorens, T.H.H., Neville, M.D.C., Sanghvi, R., Sanders, M.A., Oliver,
549 T.R.W., Leongamornlert, D., Ellis, P., Noorani, A., et al. (2021). The mutational landscape of
550 human somatic and germline cells. Nature *597*, 381–386.
551 Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016).
552 BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-
553 generation sequencing data. Bioinformatics *32*, 1749–1751.
554 Papagiannouli, F., and Mechler, B.M. (2009). discs large regulates somatic cyst cell survival and

555 expansion in Drosophila testis. Cell Res. *19*, 1139–1149.
556 Parkin, D.M., Boyd, L., and Walker, L.C. (2011). 16. The fraction of cancer attributable to
557 lifestyle and environmental factors in the UK in 2010. Br. J. Cancer *105*, S77–S81.
558 Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of
559 single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.
560 Schumacher, J., and Herlyn, H. (2018). Correlates of evolutionary rates in the murine sperm
561 proteome. BMC Evol. Biol. *18*, 35.
562 Stanley, C.E.J., and Kulathinal, R.J. (2016). flyDIVaS: A Comparative Genomics Resource for
563 Drosophila Divergence and Selection. G3 (Bethesda). *6*, 2355–2363.
564 Su, T.T., Parry, D.H., Donahoe, B., Chien, C.T., O'Farrell, P.H., and Purdy, A. (2001). Cell
565 cycle roles for two 14-3-3 proteins during Drosophila development. J. Cell Sci. *114*, 3445–3454.
566 Svetec, N., Cridland, J.M., Zhao, L., and Begun, D.J. (2016). The Adaptive Significance of
567 Natural Genetic Variation in the DNA Damage Response of Drosophila melanogaster. PLoS
568 Genet. *12*, e1005869.
569 Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J.,
570 Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al. (2019). FlyBase 2.0: the next
571 generation. Nucleic Acids Res. *47*, D759–D765.
572 Verheijen, B.M., and van Leeuwen, F.W. (2017). Commentary: The landscape of transcription
573 errors in eukaryotic cells. Front. Genet. *8*, 219.
574 Williams, G.C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence.
575 Evolution (N. Y). *11*, 398.
576 Witt, E., Benjamin, S., Svetec, N., and Zhao, L. (2019). Testis single-cell RNA-seq reveals the
577 dynamics of de novo gene transcription and germline mutational bias in Drosophila. Elife *8*,
578 e47138.
579 Witt, E., Shao, Z., Hu, C., Krause, H.M., and Zhao, L. (2021a). Single-cell RNA-sequencing
580 reveals pre-meiotic X-chromosome dosage compensation in Drosophila testis. PLOS Genet. *17*,
581 e1009728.
582 Witt, E., Svetec, N., Benjamin, S., and Zhao, L. (2021b). Transcription Factors Drive Opposite
583 Relationships between Gene Age and Tissue Specificity in Male and Female Drosophila Gonads.
584 Mol. Biol. Evol. *38*, 2104–2115.
585 Xia, B., and Yanai, I. (2022). Gene expression levels modulate germline mutation rates through

586 the compound effects of transcription-coupled repair and damage. Hum. Genet. *141*, 1211–1222.
587 Xia, B., Baron, M., Yan, Y., Wagner, F., Kim, S.Y., Keefe, D.L., Alukal, J.P., Boeke, J.D., and
588 Yanai, I. (2020). Widespread transcriptional scanning in testes modulates gene evolution rates.
589 Cell 248–262.
590 Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B.,
591 Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional
592 profiling of single cells. Nat. Commun. *8*, 14049.
593