Transcriptional and mutational signatures of the aging germline

Evan Witt, Christopher B Langer, 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 which is accompanied by changes in gene expression and mutational load. In many species including humans, old fathers pass on more paternally-derived *de novo* mutations, however, the cellular basis and cell types driving this pattern are still unclear. To understand 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 tissue from the same flies. We found that early germ cells from old and young flies have similar mutational loads, but older flies are less able to remove mutations during spermatogenesis. This indicates that germline mutations arise from primarily non-replicative factors, and that the increased mutational load of older males is due to differences in genome maintenance activities such as repairs to DNA damage. We also found that T>A mutations are enriched in older flies, and transcription-related enrichment terms are depleted in older males. Early spermatogenesis-enriched genes have lower dN/dS than late spermatogenesis-enriched genes, supporting the hypothesis that late spermatogenesis is the source of evolutionary innovation. This transcriptional disruption is reflected in the decreased expression of genome maintenance genes in early germ cells of older flies, as well as potentially aberrant transcription of transposable elements in the aging germline. Our results provide novel insights into the transcriptional and mutational signatures of the male germline.

Introduction

Aging is a process that is accompanied with complex 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 by passing a higher amount of *de novo* mutations to the offspring. These mutations mostly occur in the ovary and

testis. 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 stem cells undergo throughout the life of the male (Drost and Lee, 1995; Gao et al., 2011; Li et al., 1996). Conversely, other reports have found that the excess cell divisions do not track the ratio of maternal to paternal mutations during aging (Gao et al., 2019; Huttley et al., 2000), suggesting instead that lifestyle, chemical and environmental factors cause this discrepancy (Irigaray et al., 2007; Parkin et al., 2011). Previous studies of the effect of age on paternally inherited mutations have inferred *de novo* mutations through sequencing of parents and offspring (Gao et al., 2016). These methods are highly useful, but they only capture *de novo* mutations that have evaded repair mechanisms, ended up inside a viable gamete, fertilized an egg and created a viable embryo. Much less is known about the dynamics of mutation and repair inside the male germline. One study found that mutations arise least frequently in human spermatogonia (Moore et al., 2021). In our previous work (Witt et al., 2019), however, we instead found that mutational load is highest in the earliest stages of spermatogenesis. Taken together, these results imply that most mutations occur prior to GSC differentiation and are removed during spermatogenesis. But are these mutations replicative in origin? If so, we would expect germline stem cells from older flies to be more mutated than those from younger flies.

In addition to these mutational effects, aging is known to cause other germline phenotypes such as lower numbers of germ cells and reduced germline stem cell proliferative capacity (Lee et al., 2020). The GSC microenvironment also undergoes chemical changes associated with reductions in fecundity (Jones, 2007). These phenotypic consequences can be linked to mutation, as the germline mutational rate in young adults correlates with longevity (Cawthon et al., 2020). As such, the germline mutation rate has consequences for both an organism and its descendants.

In our previous study (Witt et al., 2019), we used single-cell RNA-sequencing (scRNA-seq) to follow germline mutations throughout *Drosophila* spermatogenesis and found evidence that germline mutations decline in abundance throughout spermatogenesis. We also found evidence of germline genome maintenance genes upregulated in germline stem cells (GSCs) and early spermatogonia, the earliest male germ cells. Our results were in line with the idea that active DNA repair plays a role in the male germline (Xia et al., 2020), but since age is an

important factor for mutational load, in this study we asked whether DNA damage repair is equally effective in older flies as in younger flies.

To study germline repair 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 mutation found 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. Additionally, we found enrichment of T>A SNPs in older flies. Hypothesizing that these mutations were due to decreased activity of genome maintenance genes, we found that genes of all types, including genome maintenance genes, are downregulated in the early germ cells of older flies. A Gene Ontology (GO) analysis confirmed that genes enriched in young early germ cells are enriched for RNA polymerase II-related functions, and several of these genes are also involved in DNA replication. We also noted several transposable elements upregulated in germ cells from older flies. Additionally, we found evidence that genes enriched in older germ cells have lower dN/dS than genes enriched in younger germ cells. Taken together, our results indicate that the old germline is highly mutated and transcriptionally dysregulated. We propose that the impaired mutational repair of the older male germline could be due to deficiencies in transcription-coupled repair (Deger et al., 2019). We also found that while the old germline is more mutated in general, early germ cells from older flies are similarly mutated as early germ cell from young flies, strengthening the notion that the increased mutational burden of the male germline is due to non-replicative factors.

Results

A cell atlas of the aged male germline

We aimed to capture representatives 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). Each of the four 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. Dotplots showing expression of key marker genes are shown in Supplemental Figure 1. In order 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, which serve as control.

Using Seurat 4 (Satija et al., 2015), we classified somatic cells into 3 broad types: hub cells, cyst cells, and epithelial cells. We split germ cells into 6 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 13,904 cells from young flies (Figure 1C) and 13,089 cells from old flies. We found that for each age group, the two replicates largely corroborate each other, with Pearson's r values over 0.91 between replicates of the same age (Supplemental Figure 2). After cell type assignments, we used Seurat 4 to perform downstream analyses on the integrated dataset.

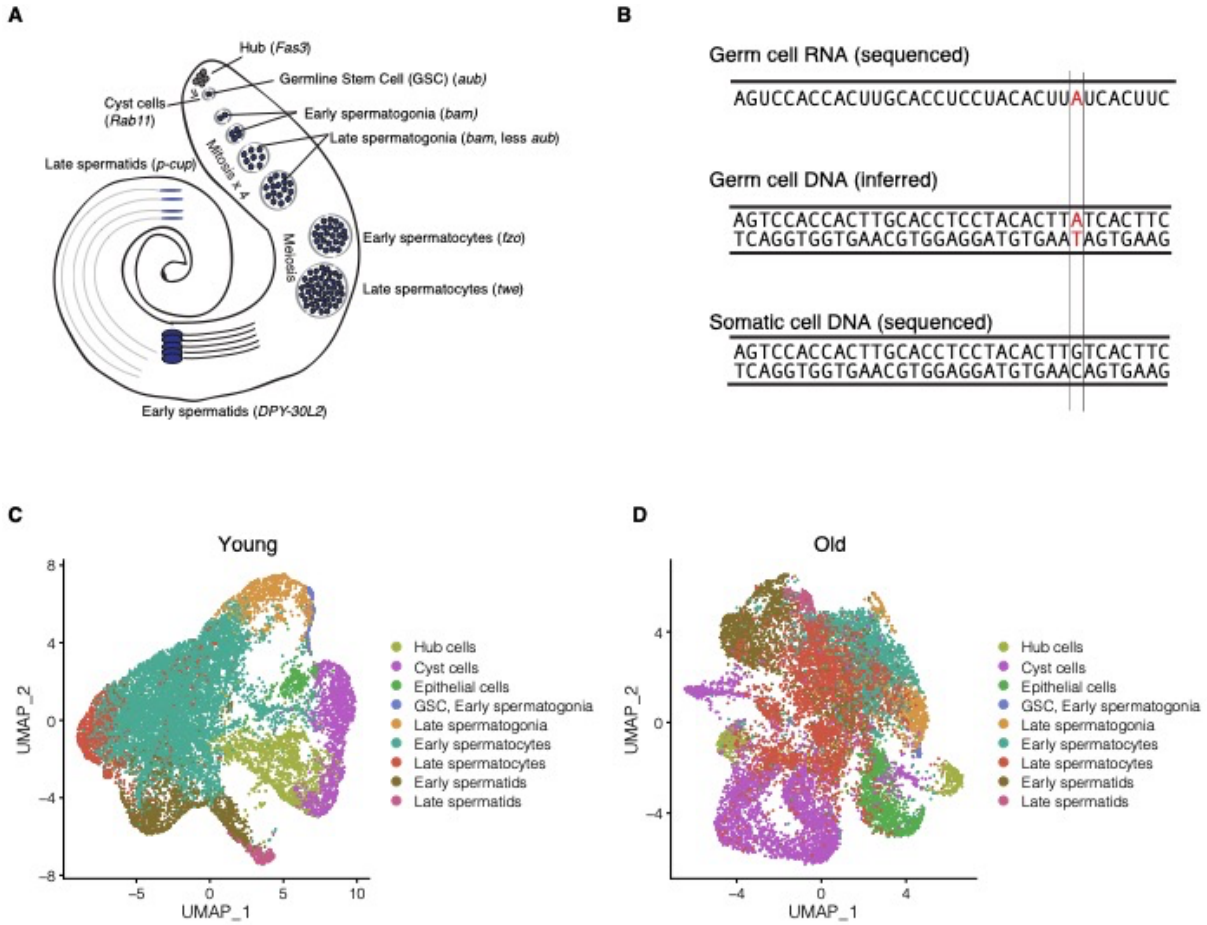


Figure 1: Overview of experimental design and visualization of old and young datasets

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

Older flies show impaired mutational repair during spermatogenesis

We identified germline SNPs in each sample and matched them to every cell with reads corroborating a given SNP. To assess mutational burden between young and old flies, we compared, for each cell type, the proportion of cells where at least one mutation was detected. In young flies, the proportion of mutated cells declines during spermatogenesis, indicating that lesions are either repaired, or mutated cells are removed from the population. Old flies begin spermatogenesis with a similar proportion of mutated GSC/early spermatogonia, but their mutational burden remains high throughout spermatogenesis (Figure 2). Proportions of mutated cells are statistically similar for young and old flies in GSC/Early spermatogonia but begin to diverge in subsequent cell types. Whereas around 50 percent of young and old GSC/early spermatogonia harbors at least one *de novo* mutation, only 8 percent of young late spermatids are mutated, compared with 40 percent of old late spermatids (Bonferroni-corrected p value = 1.25e-14). The observation that GSC/early spermatogonia has a similar number of mutated cells in young and old flies suggest that replication-related errors are not likely to be a major driver of mutations.

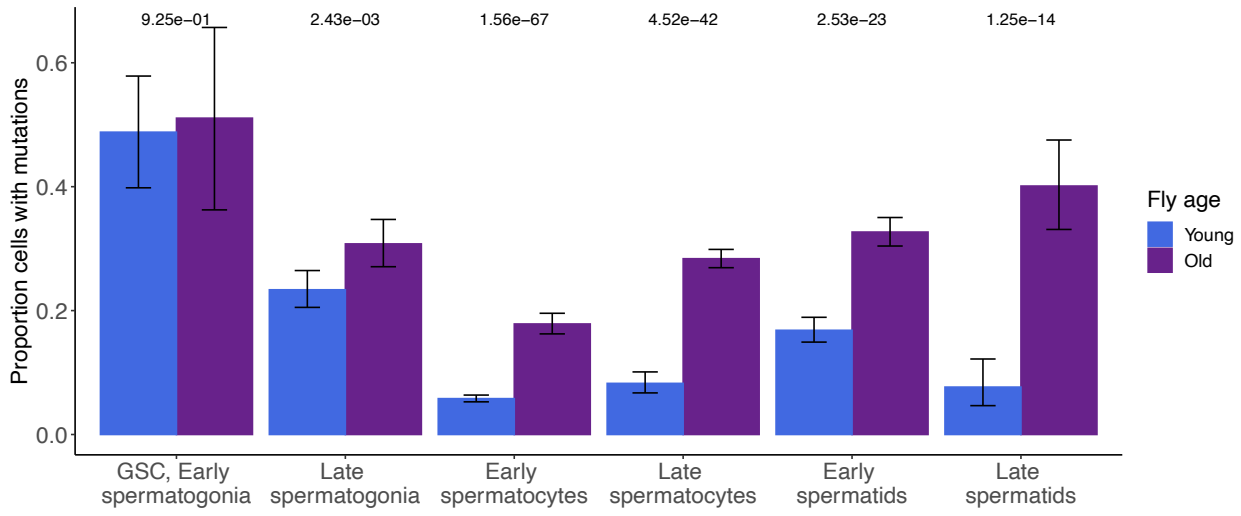


Figure 2: Old flies show impaired germline mutation repair. For old and young flies and every cell type, shown are the proportions of cells of each type carrying at least one mutation. Error bars are 95 percent confidence intervals from a chi-square test of proportions. P values Bonferroni-corrected from a chi-square test of proportions comparing young and old cells of a type. The proportion of mutated cells declines for young flies, but not old flies, during spermatogenesis.

T>A substitutions are enriched for old flies

We compared the relative proportions of the 6 major classes of mutation between young and old flies. In young flies, we found that T>G and T>C mutations are enriched compared to old flies (Figure 3A). Using a chi-square test of proportions, we found that old flies were significantly enriched for T>A transitions (Figure 3A). This suggests that there is an age-related mutational bias or repair bias in the spermatogenesis. We asked whether these mutational signatures were due to differential activity of genome maintenance genes, and found that genome maintenance genes are generally less expressed in the earliest germ cells of old flies (Figure 3B, Wilcoxon test, Bonferroni-adjusted p value = 3.04E-09). Although genome maintenance genes often perform complex functions in the cells, one possibility is that their low activity in old flies may impact the mutational bias that we observed here.

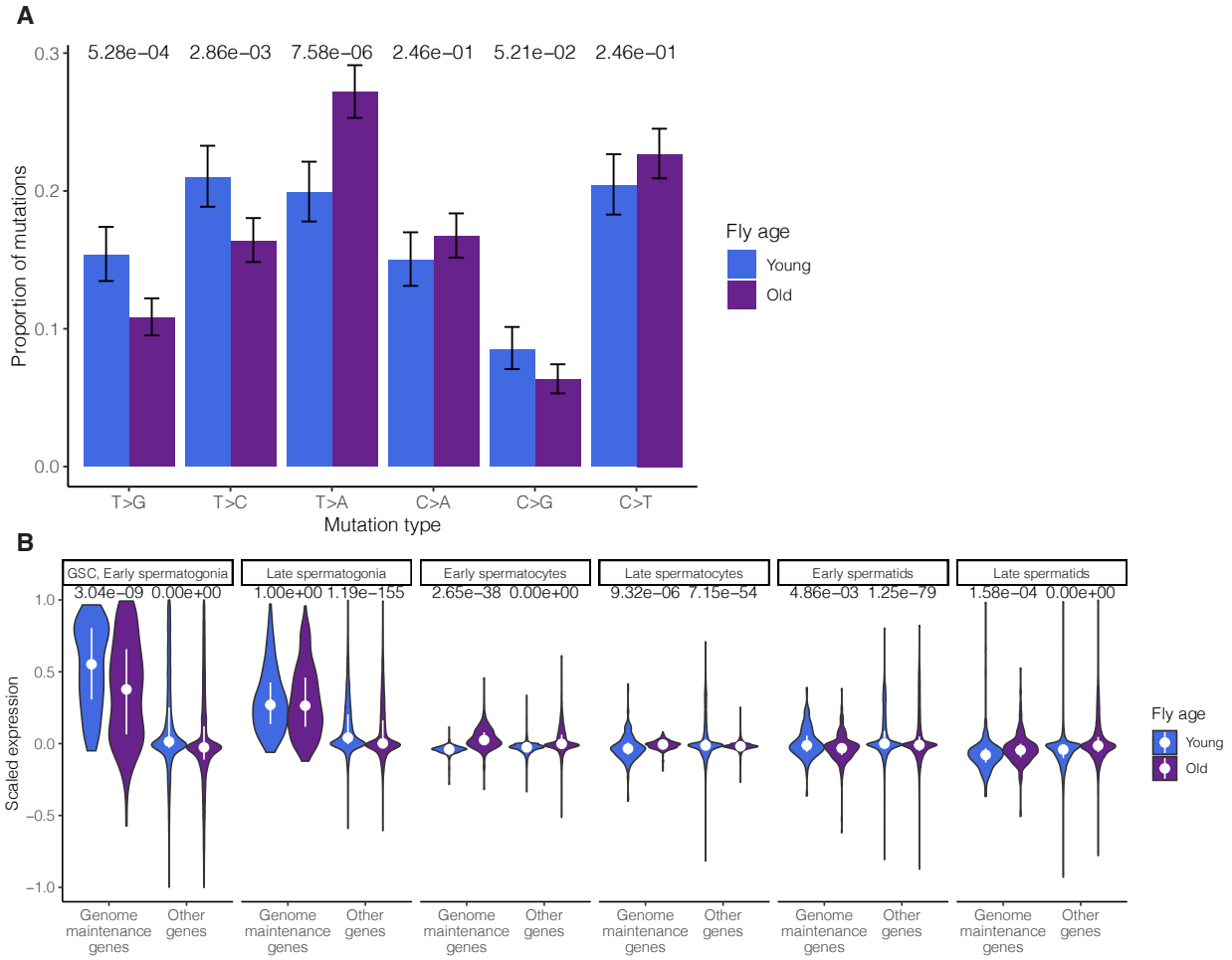


Figure 3: Age-related trends in mutational signatures and genome maintenance gene expression. A) For young and old flies, shown are the relative proportions of the 6 types of mutations (each class is equivalent to a complementary mutation, for example T>G also represents A>C). P values are Bonferroni-corrected p values of a chi-square test of proportions. T>A mutations are enriched in old flies, while T>G and T>C mutations are underrepresented. B) Scaled expression of genome maintenance genes compared to scaled expression of all other genes. 0 represents a gene's mean expression across all cell types, and the numbers on the Y axis represent standard deviations of a gene's expression in a cell type compared to its mean expression. By this measure, genome maintenance genes are less upregulated in the GSC/early spermatogonia of older flies. Additionally, other genes are also downregulated in GSC/early spermatogonia and upregulated in late spermatids, where transcription is normally suppressed. P values are Bonferroni-corrected p values of paired Wilcoxon tests between old and young flies for each group of genes. White dots are medians and vertical white lines are interquartile ranges.

Genome maintenance genes are upregulated in early germ cells from young flies

We performed differential expression testing on a list of 211 genes related to DNA damage repair compiled from our previous work (Svetec et al., 2016)(Witt et al., 2019). We

found that in GSC/early spermatogonia, 24 DNA damage response genes were significantly enriched in young flies, versus just 4 for old flies (Figure 4). In spermatocytes and spermatids, comparatively few genes are enriched in old or young flies. This corroborates our earlier observation that genome maintenance genes are generally less expressed in old GSC/early spermatogonia, which is also the most mutated cell type in both old and young flies. Depleted expression of genome maintenance genes in the earliest germ cells could impact the efficiency of germline DNA repair throughout spermatogenesis.

Many of the genome maintenance genes enriched in young GSC/early spermatogonia are RNA polymerase subunits or otherwise involved in transcription. This suggests that our observed mutational signatures of older flies might be caused by defects in transcription-coupled repair, although it is unclear how common such repair in spermatogenesis. Indeed, transcription of all genes is generally downregulated in the GSC/early spermatogonia of older flies (Figure 3B).

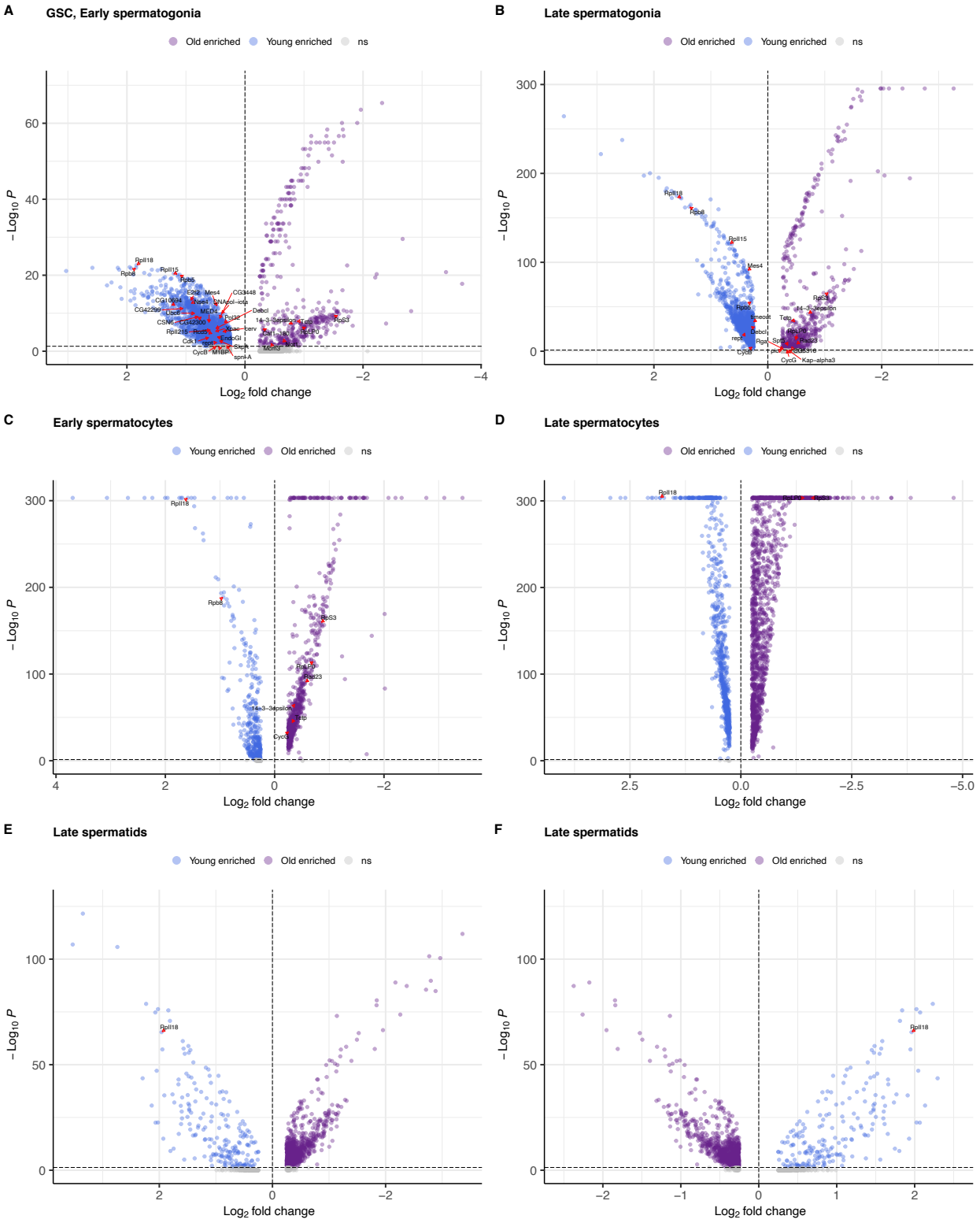


Figure 4: Age-related differential expression of genome maintenance genes. Shown are the results of differential expression tests between old and young flies, calculated separately for each cell type. Each labelled point is a gene

involved in genome maintenance with a Bonferroni adjusted P value <0.05. Many DNA repair genes are enriched for young flies in GSC, early spermatogonia, but as spermatogenesis progresses, fewer DNA repair genes are differentially expressed between old and young flies. Fold changes refer to the ratio between expression in young compared to old flies. Enrichment statistics for genome maintenance genes are in Supplemental Table 1.

Transcription-related Gene Ontology terms are enriched in early germ cells of young flies

We asked how many genes were downregulated in the germ cells of old flies, and compared the numbers of enriched genes in old and young cells of the same type for every cell type (Table 1). GSC/Early spermatogonia from young flies had 1629 genes enriched compared to old flies, whereas 254 genes were enriched in old cells compared to young. Our observed mutational phenotypes may have something to do with the substantial number of downregulated genes.

We used a PANTHER (Thomas et al., 2003) overrepresentation test to identify enriched Gene Ontology terms for differentially expressed genes between old and young flies. The top enriched term in GSC/Early spermatogonia of young flies was "transcription by RNA polymerase II" (Fold Enrichment 5.57, Bonferroni-corrected p value 3.94e-09). In contrast, GSC/Early spermatogonia of old flies showed an enrichment for "positive regulation of NIK/NF-kappaß signaling" (fold enrichment > 100, Bonferroni-corrected p value 3.49e-04). NF-kappaß is known to regulate transcription factors important for spermatogenesis and stem cell differentiation, so its misexpression could have far-reaching effects (Kaltschmidt et al., 2021; Lilienbaum et al., 2000; Pentikäinen et al., 2002; Teng et al., 2012). Late spermatids from young flies do not show enrichment of RNA polymerase II transcription and are instead enriched for tRNA transcription by RNA polymerase III (fold enrichment 80.77, adjusted p value 1.57e-03) and transcription by RNA polymerase I (fold enrichment 57.01, adjusted p value 5.10e-03). Late spermatids from old flies, on the other hand, are enriched for cytoplasmic translation (fold enrichment 13.9, adjusted p value 8.96e-55) and ribosomal small subunit assembly (fold enrichment 8.43, adjusted p value 8.73e-03). These results support our earlier findings that transcription is globally downregulated in the GSC/early spermatogonia of old flies.

Cell type	# genes enriched in young flies	# genes enriched in old flies
GSC, Early spermatogonia	1629	254
Late spermatogonia	460	450
Early spermatocytes	329	741
Late spermatocytes	665	1483
Early spermatids	449	471
Late spermatids	153	787

Table 1: Numbers of age-enriched genes per cell type. Arranged from top (early) to bottom (latest). For young flies, the cell type with the most enriched genes is GSC/Early spermatogonia, the earliest germ cell class. Late spermatids are the most mutated class of cell in older flies. In every class of cells except GSC, early and late spermatogonia, old flies have more enriched genes than young flies.

Old flies show cell-type specific enrichment of *de novo* genes

Our previous observation that transcription is broadly downregulated in early germ cells led us to ask whether these trends differentially impact special classes of transcripts, including transposable elements and *de novo* genes (genes born from previously non-genic DNA) (Reinhardt et al., 2013; Schlotterer, 2015; Zhao et al., 2014; Zhou et al., 2008). We constructed an alternate Cellranger reference containing segregating and fixed *de novo* genes characterized in our previous work (Witt et al., 2019), as well as transposable elements (TEs) from another study (Lawlor et al., 2021). We observed that as a class, scaled expression of *de novo* genes and TEs follows patterns of other genes in older flies, indicating that they are also subject to the effects of age-related dysregulation (Supplemental Figure 3). Segregating and fixed *de novo* genes are depleted in early germ cells of old flies and are enriched in late spermatids from old flies, similar to other genes. While most *de novo* genes were not enriched or depleted between old and young flies, 5 fixed and one segregating *de novo* gene were enriched in cells from old flies, whereas 1 fixed and 3 segregating *de novo* genes were enriched in young flies (Figure 5).

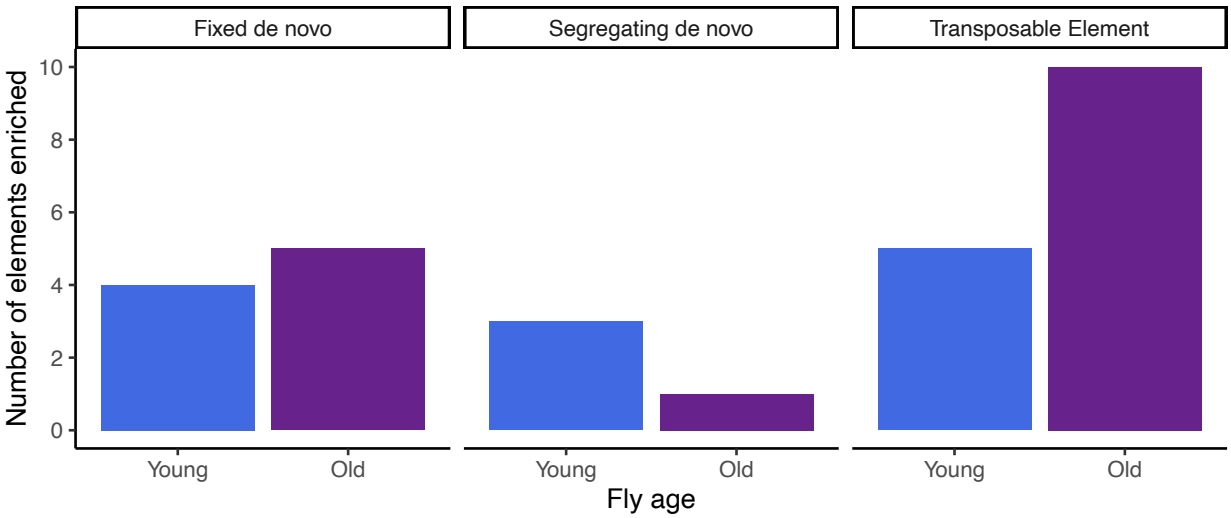


Figure 5: Enrichment of other sequence elements in old flies. Shown are the numbers of transposable elements and *de novo* genes differentially expressed between young and old flies. Fixed *de novo* genes are similarly likely to be enriched in either old or young flies, but segregating *de novo* genes are more commonly enriched in young flies. Transposable elements are more likely to be enriched in old flies. Enrichment statistics for each gene and TE are in Supplemental Tables 2 and 3.

Old flies show cell-type specific enrichment of several transposable elements

One hypothesis associated with TEs and aging is that reduced chromatin silencing in old flies may increase the expression of TE (Wood et al., 2016). We identified sequence elements enriched between old and young flies for each cell type, and found that 10 transposable elements are enriched in old flies, versus 5 in young flies (Figure 5). Most notably, Jockey2 (Tambones et al., 2019) was enriched in cyst, hub and epithelial cells as well as early and late spermatocytes and spermatids. Most transposable elements are highly expressed in old than young flies, however, the fold changes are moderate, with most showing differences smaller than 2-fold, suggesting that TE probably plays a minor role in aging testis compared to other tissues (Li et al., 2013). The increased activation of transposable elements in old flies could be a consequence of transcriptional dysregulation or reduced chromatin silencing, as most of these TEs likely are inhibited in spermatogenesis in young flies. It would be interesting to study this in detail and to understand the impact of TE in changes of gene expression in aging testis.

Early spermatogenesis-enriched genes have lower dN/dS than late spermatogenesis-enriched genes; young-enriched genes have higher dN/dS than old-enriched genes

We asked whether functional constraint varies for age-specific or cell-type specific genes. We defined "age-enriched" genes as genes differentially expressed between old and young flies in the same cell type. We defined "cell-type-enriched" genes as genes differentially expressed between cell types of the same age. To find cell type-enriched 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, we then compared dN/dS values for early-enriched and late-enriched genes. In both young and old flies, late-enriched genes have higher dN/dS values than genes enriched in GSC/early spermatogonia (Figure 6A), in line with the idea that genes expressed in late spermatogenesis may evolve rapidly. Note that spermatocytes and spermatids are also hotspot 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-specific genes, we subset 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 enriched in young cells have higher dN/dS than genes enriched in old cells (Figure 6B), indicating that genes enriched in the young male germline are less mutationally constrained compared to genes enriched in the older germline.

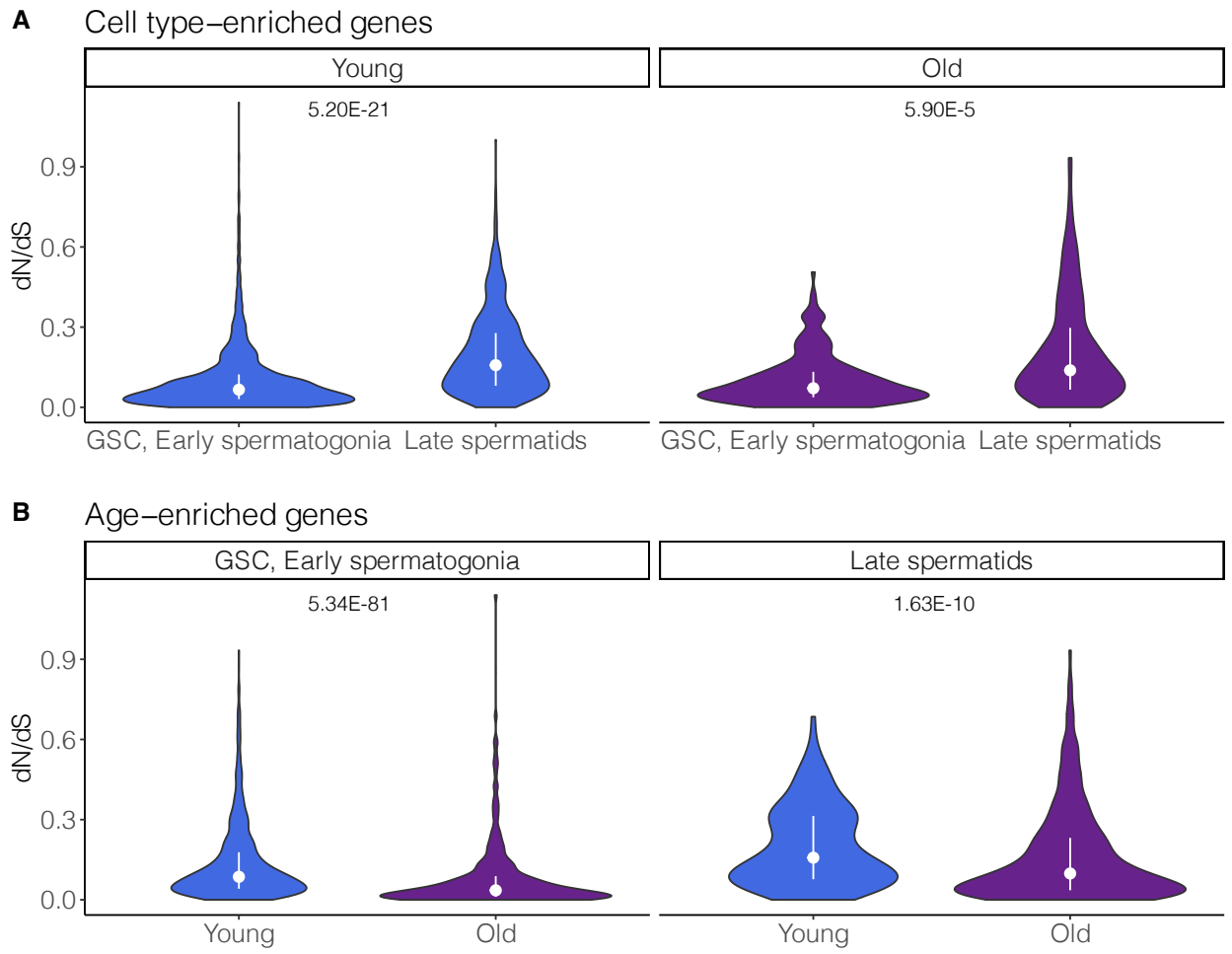


Figure 6: dN/dS trends of cell type-enriched and age-enriched genes. We calculated gene enrichments in two different ways. First, we calculated gene enrichment between GSC/early spermatogonia and late spermatids separately for young and old flies. Then, we identified genes enriched in young and old cells 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 p values 2.6E-21, 5.9E-5, respectively). B) Age-enriched genes: In GSC/early spermatogonia and late spermatids, young-enriched genes have higher dN/dS than old-enriched genes (Wilcoxon p values 2.67E-81 and 8.16E-11, respectively). Figure shows Bonferroni-corrected p values.

Discussion

Mutational load is an equilibrium between mutation and repair, and in old flies, this equilibrium tips 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. This finding adds a new explanation for the still-controversial mechanism behind the increased age-dependent mutational load of the male germline. Our findings suggest that rather than slowly accumulating

mutations in cycling germ cells, the older germline starts with a similar mutational burden to younger flies but lacks the ability to repair the mutations or remove the cells carrying them. Our work corroborates previous work that found that the huge excess of male germline divisions is too large to explain the much smaller ratio of male/female-inherited mutations during parental aging (Gao et al., 2019). Our finding that early germ cells from young and old flies are similarly mutated supports the notion that most germline mutations are not due to replicative processes. Some of our conclusions are in line with recent findings that germline stem cells have the lowest mutation rate of any human cell type (Moore et al., 2021).

In addition to being highly mutated, the older germline shows distinct mutational signatures compared to the younger germline. We found a statistical overrepresentation of T>A mutations in old flies, and an underrepresentation of T>G and T>C mutations. These altered ratios of single nucleotide polymorphisms could be caused by differential activity of DNA repair pathways in the old germline. Indeed, we found differential expression of genome maintenance genes between young and old GSC/early spermatogonia, the cell type with the highest expression of genome maintenance genes. Altered types and numbers of *de novo* mutations would likely have implications for a population, affecting the type and frequency of genetic novelties that emerge (Loewe and Hill, 2010). In the future, it would be interesting to understand the molecular mechanisms contributing to mutational bias in germ cells.

The depleted genome maintenance genes in the old early germline are likely a consequence of globally diminished transcription. We observed that scaled expression of all genes is generally down in GSC/early spermatogonia, but up in late spermatids. The latter result is intriguing because transcription largely ceases after meiosis in the male Drosophila germline (Barreau et al., 2008). While the downregulation of transcription in early germ cells could have implications for germline DNA repair, the potential effects of increased post-meiotic transcription are less clear. It could have no effect, or it could affect spermatid maturation or sperm competition, potentially affecting fertility. Indeed, increased male age associates with reduced fertility in humans (Harris et al., 2011).

Other studies have proposed that the testis uses ubiquitous gene expression to detect genomic lesions and repair them with transcription-coupled repair (Xia and Yanai, 2021; Xia et al., 2020). Due to our bias towards detecting mutations in expressed genes, our data is not ideal to test this hypothesis. We noted, however, that genes with many detectable SNPs tend to be

lowly expressed across replicates (Supplemental Figure 4), a finding that appears to support the transcriptional scanning model.

The global post-meiotic upregulation of transcription extends beyond conserved genes. We observed that many transposable elements are significantly enriched in old germ cells, and transcription of transposable elements is generally upregulated. Not only could this have implications for transposable element mobilization, it could also create heritable changes in chromatin structure, signaling, or gene expression (Chuong et al., 2017; Lanciano and Cristofari, 2020).

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

Unexpectedly, we also found that, within analogous cell types, genes enriched in cells from young flies have higher dN/dS than genes enriched in old flies. Rapidly evolving genes may provide a greater evolutionary advantage to young flies than old flies. Our results are interesting in the light of recent work which found that genes expressed later in life fix nonsynonymous mutations more frequently (Cheng and Kirkpatrick, 2021). One should note that their methodology is different: their age-biased genes were identified from whole-body data whereas ours were calculated just from germ cells. Gene expression in the testis is often an outlier compared to other tissues (Witt et al., 2021b), so the results of these two studies are not necessarily in conflict. Nevertheless, the consistent pattern between this study and that of Cheng and Kirkpatrick is that genes enriched in late spermatids have a higher dN/dS than those enriched in early germ cells. In this way, developmental trends of molecular evolution are similar between whole-organism development and germline development.

Our study design is limited because it detects mutations in expressed transcripts. While we have strict criteria on the identification of novel SNPs, the abundance of false negatives could vary between cell types due to cell-specific variation in transcriptional activity. We are reassured because the most commonly mutated cell type in our datasets is GSC/early spermatogonia,

consistent with our previous observations. If transcriptional activity biased our inference of mutational load we would expect spermatocytes, the most transcriptionally active cell type, to appear the most mutated instead. This potential confounder would be resolved if a method became available to simultaneously perform RNA-seq and whole-genome sequencing on single cells.

Another potential confounder is that aging might create the appearance of germline SNPs through reduced transcriptional fidelity (Verheijen and van Leeuwen, 2017). We do not think this is a significant source of error, since our SNPs are verified by multiple independent reads and cannot be present in more than one dataset. Furthermore, if aging cells were consistently producing more inaccurate transcript than their younger counterparts, we would have found that GSC/early spermatogonia appeared more commonly mutated in old flies than young. Instead, we found a statistically similar proportion of GSC/early spermatogonia with mutations in both age groups, with the divergence between the two appearing in late spermatogonia and increasing during spermatogenesis. While it is unlikely that age-related transcriptional fidelity significantly impacts our results, this topic would benefit from combined scRNA/DNA sequencing from the same cells. The technology allowing us to trace *de novo* mutations throughout the germline is still very new, and we look forward to technological advancements in this exciting new field.

Methods

ScRNA-seq library preparation and sequencing

We dissected testis that were 48 hours old or 25 days old. Single-cell suspensions for testis samples were prepared as described in our previous work (Witt et al., 2021a). Libraries were prepared with 10X Chromium 3' V3 kit and sequenced with Illumina Hiseq 4000.

Genomic DNA preparation

Fly carcasses were frozen at -80°C, then ground in 200 µL 100 mM Tris-HCl, pH 7.5, 100 mM EDTA, 100mM NACl, 0.5% SDS. The mixtures were incubated at 65°C for 40 minutes. Then, 160 µl KAc and 240 µl 6M LiCl were added, tubes were inverted 10 times and placed on ice for 30 minutes. Samples were then centrifuged at 18000g at 4°C for 15 minutes. The supernatant was transferred into a new tube and an equivalent volume of isopropanol was added and mixed by inversion. Samples were spun for 15 minutes at 18000 g and the supernatant

was discarded. Pellets were washed with 800 µL 70% ethanol, samples were spun at 18000 g for 5 minutes, and the supernatant was discarded. Pellets were air dried for 5 minutes then resuspended in 100 µL nuclease-free water. DNA was then used for Illumina library preparation and sequencing by an Illumina Highseq4000 machine.

ScRNA-seq data processing

ScRNA-seq data were aligned with Cellranger Count and further processed with Seurat. To assign cell types with Seurat, we used marker genes described in Witt et al. 2021 (Witt et al., 2021a). Clustering and cell annotation was performed separately for the young and old samples, with canonical correlation analysis used to integrate libraries of each type. The young and old datasets were later integrated into a combined dataset with canonical correlation analysis to facilitate differential expression analysis.

SNPs were called with bcftools (Narasimhan et al., 2016) separately for each single-cell library and each gDNA library. Per-base coverage was calculated for every gDNA sample with Samtools (Li et al., 2009). For each young and old SC library, bcftools isec was used to extract mutations only present in the SC data and not the somatic gDNA. Using Samtools, we identified every cell barcode in the scRNA-seq data that corroborated every SNP (details in accompanying code). For each mutated position, we then verified that the corresponding locus in the gDNA file had at least 10 reads supporting the reference allele, and 0 reads supporting the alt allele. We also required that every SNP be present only in a single scRNA-seq dataset, to reduce the chance that RNA editing events or transcription errors caused us to falsely infer a SNP. We also required every SNP to have >1 read corroborating it, reducing the potential impact of sequencing errors.

Comparisons using scaled expression

To compare gene expression for groups of genes across replicates, we scaled expression using the ScaleData Seurat function separately on each replicate. Expression is scaled such that 0 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 test and p values were adjusted with Bonferroni's correction.

Figures were made with ggplot2 (Gómez-Rubio, 2017) and statistics were performed with the rstatix package.

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 an adjusted p value < 0.05 and fold change > 0 as enriched in young, and <0 as enriched in old. We then constructed volcano plots with the EnhancedVolcano package, labeling differentially expressed genes from the list of 211 genome maintenance genes from our previous studies (Svetec et al., 2016; 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 in the previous section. 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 has been deposited to NCBI BioProject PRJNA777411.

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 Scholar Program (VS-2020-35), and an Alfred P. Sloan Research Fellowship (FG-2018-10627) to L. Z.

Declaration of interests

The authors declare no competing interests.

Reference

Barreau, C., Benson, E., Gudmannsdottir, E., Newton, F., and White-Cooper, H. (2008). Post-meiotic transcription in *Drosophila* testes. Development *135*, 1897–1902.

Cawthon, R.M., Meeks, H.D., Sasani, T.A., Smith, K.R., Kerber, R.A., O'Brien, E., Baird, L., Dixon, M.M., Peiffer, A.P., Leppert, M.F., et al. (2020). Germline mutation rates in young adults predict longevity and reproductive lifespan. Sci. Rep. *10*, 10001.

Cheng, C., and Kirkpatrick, M. (2021). Molecular evolution and the decline of purifying selection with age. Nat. Commun. *12*, 2657.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

Crow, J.F. (2000). The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. *1*, 40–47.

Deger, N., Yang, Y., Lindsey-Boltz, L.A., Sancar, A., and Selby, C.P. (2019). Drosophila, which lacks canonical transcription-coupled repair proteins, performs transcription-coupled repair. J. Biol. Chem. *294*, 18092–18098.

Drost, J.B., and Lee, W.R. (1995). Biological basis of germline mutation: Comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. Environ. Mol. Mutagen. *25*, 48–64.

Gao, J.-J., Pan, X.-R., Hu, J., Ma, L., Wu, J.-M., Shao, Y.-L., Barton, S.A., Woodruff, R.C., Zhang, Y.-P., and Fu, Y.-X. (2011). Highly variable recessive lethal or nearly lethal mutation rates during germ-line development of male Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *108*, 15914–15919.

Gao, Z., Wyman, M.J., Sella, G., and Przeworski, M. (2016). Interpreting the dependence of mutation rates on age and time. PLoS Biol. *14*, 1–16.

Gao, Z., Moorjani, P., Sasani, T.A., Pedersen, B.S., Quinlan, A.R., Jorde, L.B., Amster, G., and Przeworski, M. (2019). Overlooked roles of DNA damage and maternal age in generating human germline mutations. Proc. Natl. Acad. Sci. *116*, 9491 LP – 9500.

Gómez-Rubio, V. (2017). ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). J. Stat.

Softw. *77*.
Harris, I.D., Fronczak, C., Roth, L., and Meacham, R.B. (2011). Fertility and the aging male. Rev. Urol. *13*, e184–e190.
Huttley, G.A., Jakobsen, I.B., Wilson, S.R., and Easteal, S. (2000). How important is DNA replication for mutagenesis? Mol. Biol. Evol. *17*, 929–937.
Irigaray, P., Newby, J.A., Clapp, R., Hardell, L., Howard, V., Montagnier, L., Epstein, S., and Belpomme, D. (2007). Lifestyle-related factors and environmental agents causing cancer: An overview. Biomed. Pharmacother. *61*, 640–658.
Jones, D.L. (2007). Aging and the germ line: where mortality and immortality meet. Stem Cell Rev. *3*, 192–200.
Kaltschmidt, C., Greiner, J.F.W., and Kaltschmidt, B. (2021). The Transcription Factor NF-κB in Stem Cells and Development. Cells *10*, 2042.
Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. Nat. Rev. Genet. *21*, 721–736.
Lawlor, M.A., Cao, W., and Ellison, C.E. (2021). A burst of transposon expression accompanies the activation of Y chromosome fertility genes during Drosophila spermatogenesis. BioRxiv 2021.05.10.443472.
Lee, M.-H., Luo, H.-R., Bae, S.H., and San-Miguel, A. (2020). Genetic and Chemical Effects on Somatic and Germline Aging. Oxid. Med. Cell. Longev. *2020*, 4684890.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., and Dubnau, J. (2013). Activation of transposable elements during aging and neuronal decline in Drosophila. Nat. Neurosci. *16*, 529–531.
Li, W.H., Ellsworth, D.L., Krushkal, J., Chang, B.H., and Hewett-Emmett, D. (1996). Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol. Phylogenet. Evol. *5*, 182–187.
Lilienbaum, A., Sage, J., Mémet, S., Rassoulzadegan, M., Cuzin, F., and Israël, A. (2000). NF-kappa B is developmentally regulated during spermatogenesis in mice. Dev. Dyn. an Off. Publ. Am. Assoc. Anat. *219*, 333–340.

Loewe, L., and Hill, W.G. (2010). The population genetics of mutations : good , bad and indifferent. 1153–1167.

Moore, L., Cagan, A., Coorens, T.H.H., Neville, M.D.C., Sanghvi, R., Sanders, M.A., Oliver, T.R.W., Leongamornlert, D., Ellis, P., Noorani, A., et al. (2021). The mutational landscape of human somatic and germline cells. Nature.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics *32*, 1749–1751.

Parkin, D.M., Boyd, L., and Walker, L.C. (2011). 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br. J. Cancer *105*, S77–S81.

Pentikäinen, V., Suomalainen, L., Erkkilä, K., Martelin, E., Parvinen, M., Pentikäinen, M.O., and Dunkel, L. (2002). Nuclear factor-kappa B activation in human testicular apoptosis. Am. J. Pathol. *160*, 205–218.

Reinhardt, J.A., Wanjiru, B.M., Brant, A.T., Saelao, P., Begun, D.J., and Jones, C.D. (2013). De novo ORFs in Drosophila are important to organismal fitness and evolved rapidly from previously non-coding sequences. PLoS Genet. *9*, e1003860.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schlotterer, C. (2015). Genes from scratch - the evolutionary fate of de novo genes. Trends Genet. *31*, 215–219.

Schumacher, J., and Herlyn, H. (2018). Correlates of evolutionary rates in the murine sperm proteome. BMC Evol. Biol. *18*, 35.

Svetec, N., Cridland, J.M., Zhao, L., and Begun, D.J. (2016). The Adaptive Significance of Natural Genetic Variation in the DNA Damage Response of Drosophila melanogaster. PLoS Genet. *12*, e1005869.

Tambones, I.L., Haudry, A., Simão, M.C., and Carareto, C.M.A. (2019). High frequency of horizontal transfer in Jockey families (LINE order) of drosophilids. Mob. DNA *10*, 43.

Teng, Y.-N., Chuang, P.-J., and Liu, Y.-W. (2012). Nuclear factor-κB (NF-κB) regulates the expression of human testis-enriched Leucine-rich repeats and WD repeat domain containing 1 (LRWD1) gene. Int. J. Mol. Sci. *14*, 625–639.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K.,

Muruganujan, A., and Narechania, A. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Res. *13*, 2129–2141.
Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al. (2019). FlyBase 2.0: the next generation. Nucleic Acids Res. *47*, D759–D765.
Verheijen, B.M., and van Leeuwen, F.W. (2017). Commentary: The landscape of transcription errors in eukaryotic cells. Front. Genet. *8*, 219.
Witt, E., Benjamin, S., Svetec, N., and Zhao, L. (2019). Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in drosophila. Elife *8*, e47138.
Witt, E., Shao, Z., Hu, C., Krause, H.M., and Zhao, L. (2021a). Single-cell RNA-sequencing reveals pre-meiotic X-chromosome dosage compensation in Drosophila testis. PLOS Genet. *17*, e1009728.
Witt, E., Svetec, N., Benjamin, S., and Zhao, L. (2021b). Transcription Factors Drive Opposite Relationships between Gene Age and Tissue Specificity in Male and Female Drosophila Gonads. Mol. Biol. Evol. *38*, 2104–2115.
Wood, J.G., Jones, B.C., Jiang, N., Chang, C., Hosier, S., Wickremesinghe, P., Garcia, M., Hartnett, D.A., Burhenn, L., Neretti, N., et al. (2016). Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. Proc. Natl. Acad. Sci. U. S. A. *113*, 11277–11282.
Xia, B., and Yanai, I. (2021). Gene expression levels modulate germline mutation rates through the compound effects of transcription-coupled repair and damage. Hum. Genet.
Xia, B., Baron, M., Yan, Y., Wagner, F., Kim, S.Y., Keefe, D.L., Alukal, J.P., Boeke, J.D., and Yanai, I. (2020). Widespread transcriptional scanning in testes modulates gene evolution rates. Cell 248–262.
Zhao, L., Saelao, P., Jones, C.D., and Begun, D.J. (2014). Origin and spread of de novo genes in Drosophila melanogaster populations. Science (80-.). *343*, 769–772.
Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049.
Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and

Wang, W. (2008). On the origin of new genes in Drosophila. Genome Res. *18*, 1446–1455.

Supplementary information for

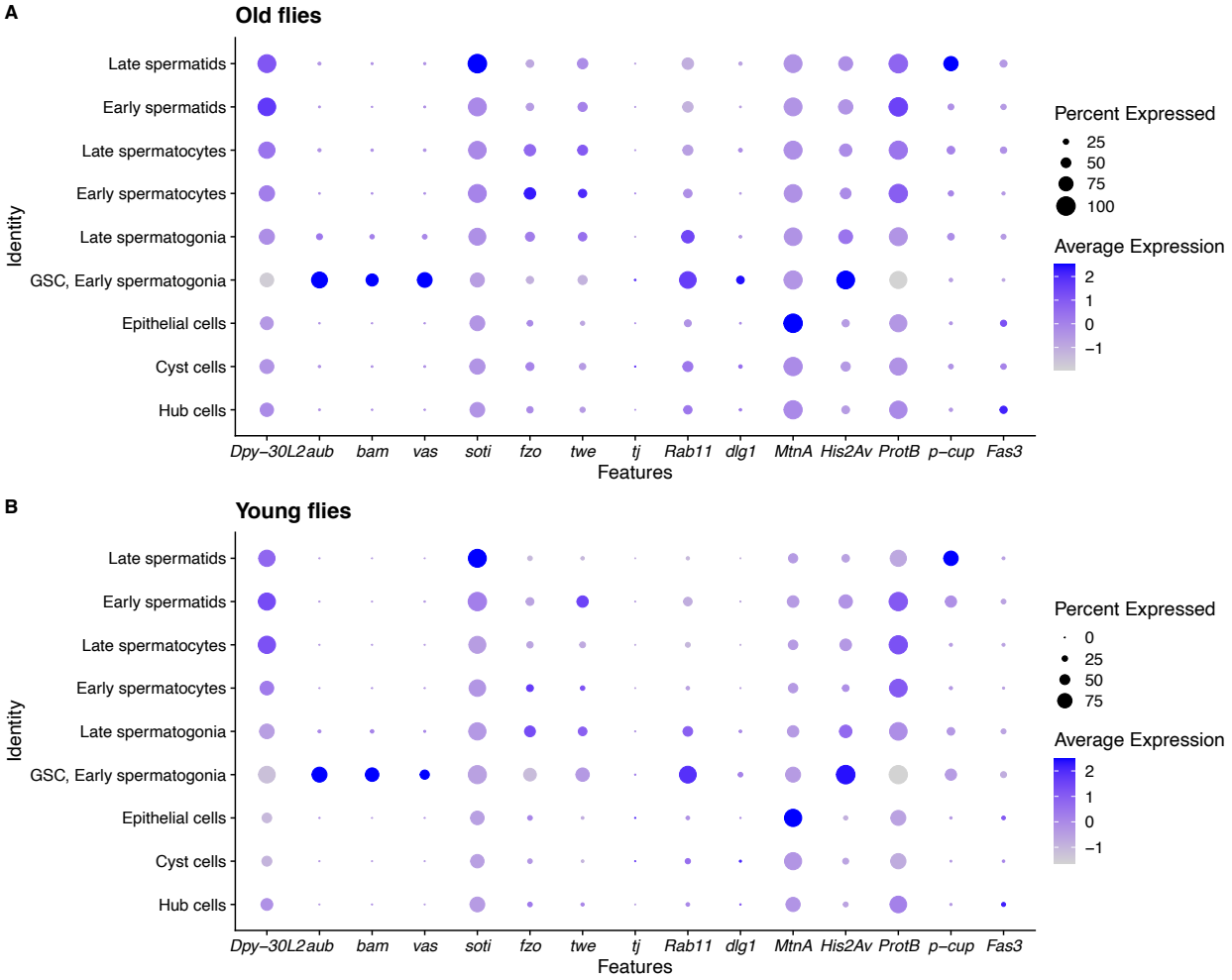
Transcriptional and mutational signatures of the aging germline

Evan Witt, Christopher B Langer, Li Zhao

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

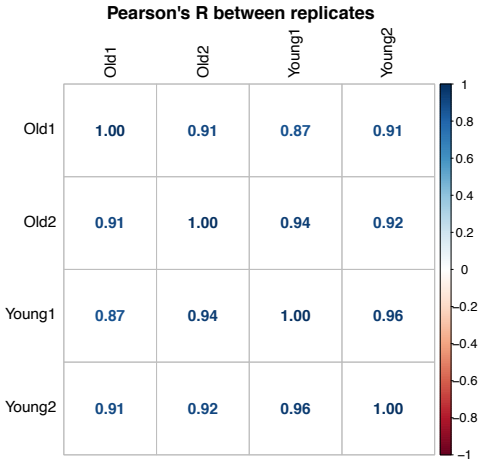
*Correspondence to: lzhao@rockefeller.edu

Supplemental figures and tables:

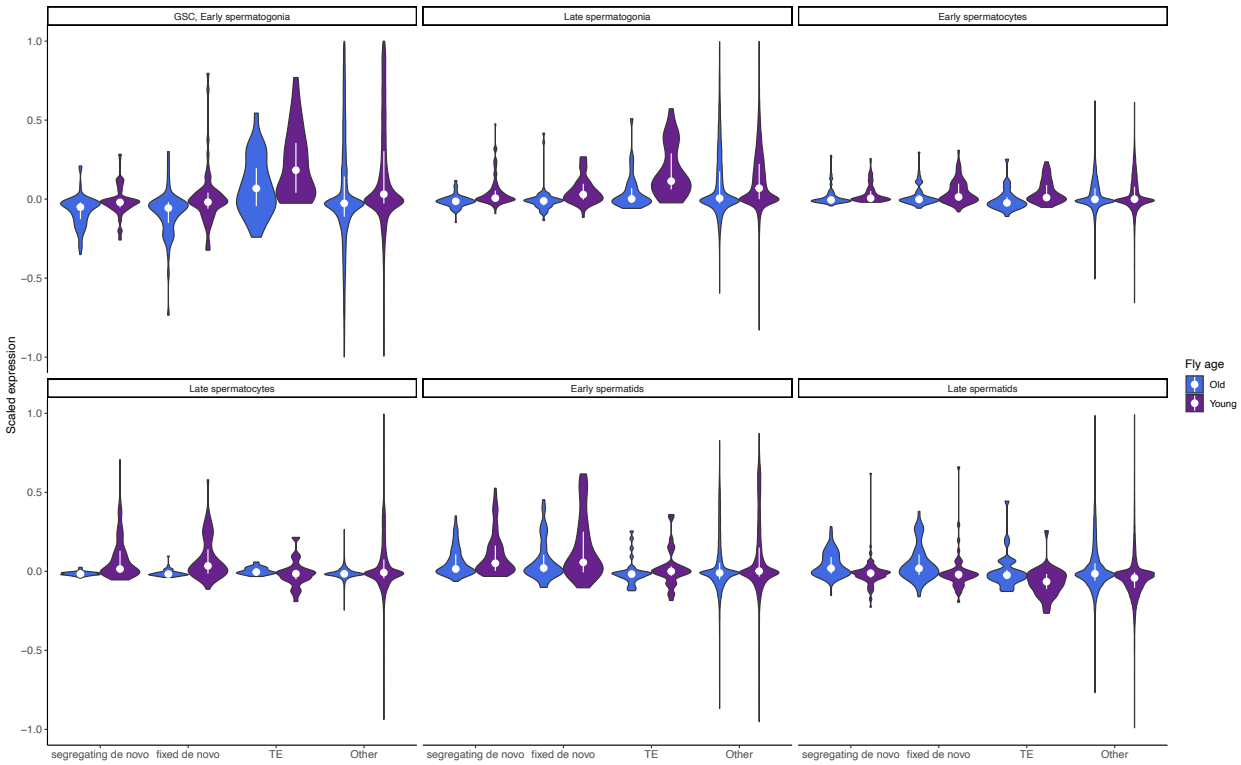


Supplemental Figure 1: Dotplots of key marker genes in old, young fly testes. Split by cell type, these are the average expression values of the "SCT" slot in the old (A) and young (B) Seurat objects. Color corresponds to the level of expression, and the size of the dot represents the percent of cells of a class where a gene is detected. For example, p-cup was used to assign cells as late spermatids. *Bam* (Ohlstein and McKearin, 1997; Witt et al., 2019), *aub* (Rojas-Ríos et al., 2017) and *vas* (Liu et al., 2009) were used to assign cells as GSC/early spermatogonia. *MtnA*(Faisal et al., 2014) was used to assign cells as epithelial cells, and *dlg1* (Papagiannouli and Mechler, 2009) was used to assign cyst cells. *Dpy-30L2* (Vardanyan et al., 2008) was used to differentiate early spermatids from late

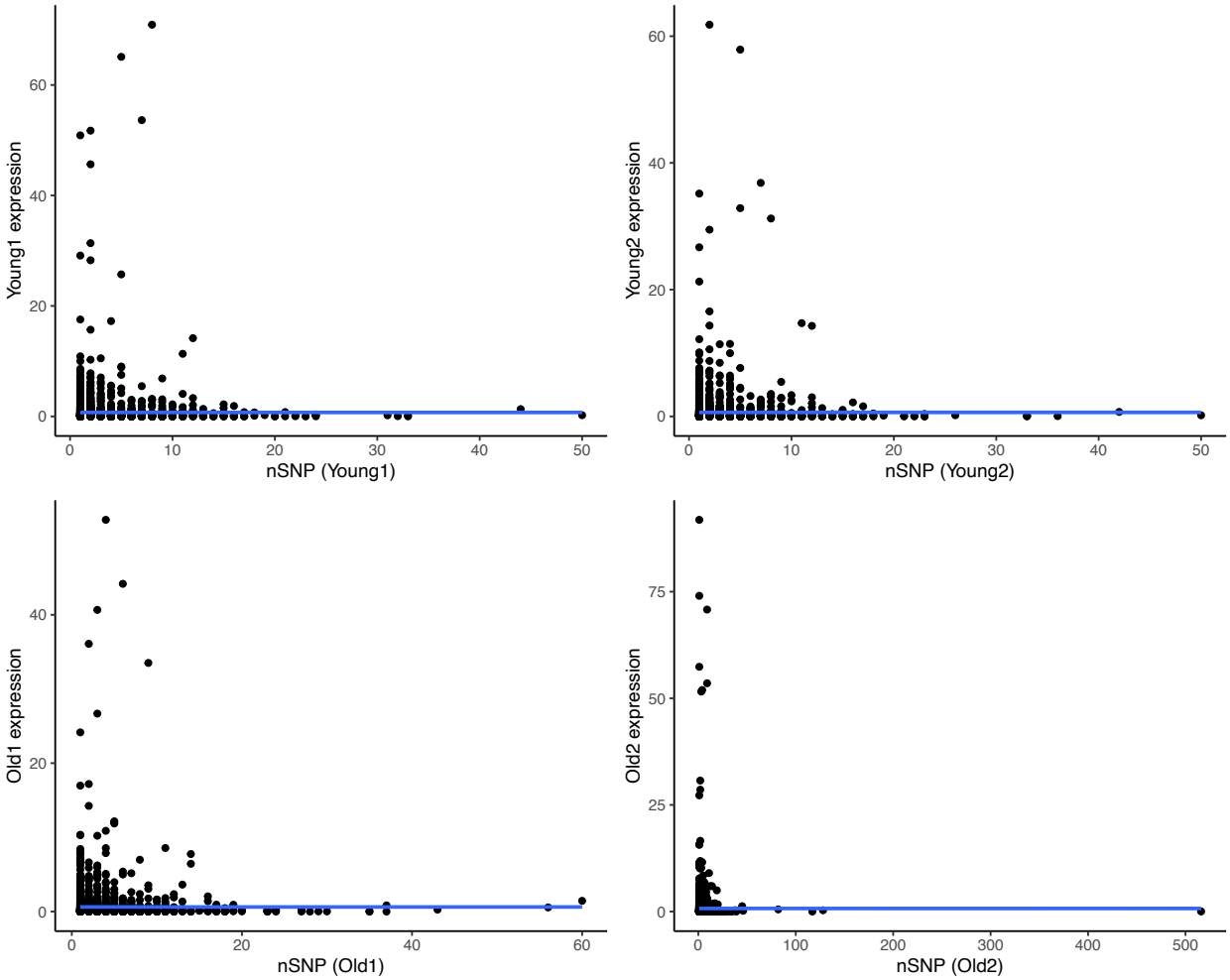
spermatocytes, and *fzo* (Hwa et al., 2002) was used to assign early spermatocytes. Late spermatogonia were assigned with *aub, bam, vas*, while having less *His2Av* (Jayaramaiah Raja and Renkawitz-Pohl, 2005) than GSC/early spermatogonia.



Supplemental Figure 2: Correlogram of scRNA-seq replicates. Replicates from each age group all correlate with Pearson's R >0.91. Correlations were drawn from gene expression values from the "RNA" slot of the Seurat object using the corrplot R package.



Supplemental Figure 3: Scaled expression of *de novo* genes and transposable elements across cell types and ages. 0 represents a gene's mean expression across all cell types and +/- 1 represent +/- 1 standard deviation from its mean expression.



Supplemental Figure 4: Expression vs number of detected SNPS for genes in every replicate.
Shown are the SNPS present in each scRNA-seq vcf file but not in the gDNA vcf file for the same sample. For each gene with detectable SNPS, shown is the number of SNPs detected in that replicate as well as the average expression value of that gene across every cell. While a Loess regression does not point to a clear trend, it is notable that genes with many SNPS tend to be lowly expressed. Intergenic SNPS are not included in this analysis.

Gene	P value	Log_2 fold change	Bonferroni-adjusted P	Fly age	Cell type
RpII18	1.75E-224	1.59384074	2.77E-220	Young	Early spermatocytes
Rpb8	2.01E-114	0.90727248	3.18E-110	Young	Early spermatocytes

RpS3	1.20E-193	0.77368825	1.90E-189	Old	Early spermatocytes
Rad23	7.10E-164	0.58562099	1.12E-159	Old	Early spermatocytes
RpLP0	3.15E-143	0.63004696	4.99E-139	Old	Early spermatocytes
CkIalpha	1.83E-122	0.3738338	2.90E-118	Old	Early spermatocytes
lwr	2.06E-116	0.39495396	3.26E-112	Old	Early spermatocytes
Ubc6	7.99E-112	0.3512995	1.26E-107	Old	Early spermatocytes
CG42300	8.21E-105	0.37298173	1.30E-100	Old	Early spermatocytes
Caf1-55	1.87E-98	0.288817	2.95E-94	Old	Early spermatocytes
CG42299	1.06E-91	0.29768116	1.67E-87	Old	Early spermatocytes
EndoGI	5.27E-88	0.26522466	8.33E-84	Old	Early spermatocytes
rept	1.46E-80	0.29615535	2.31E-76	Old	Early spermatocytes
14-3-3epsilon	3.28E-54	0.33395668	5.19E-50	Old	Early spermatocytes
Tctp	6.80E-41	0.28827023	1.08E-36	Old	Early spermatocytes
RpII18	0	1.68662125	0	Young	Early spermatids
Rpb8	5.29E-266	0.85519658	8.37E-262	Young	Early spermatids
RpS3	7.41E-41	0.47481042	1.17E-36	Old	Early spermatids
Ubc6	1.86E-39	0.33477825	2.94E-35	Old	Early spermatids
RpLP0	1.13E-35	0.42494605	1.79E-31	Old	Early spermatids
14-3-3epsilon	1.50E-30	0.29197628	2.38E-26	Old	Early spermatids
RpII18	1.13E-251	1.25704367	1.78E-247	Young	Cyst cells
Rpb8	2.07E-127	0.68485369	3.27E-123	Young	Cyst cells
RpII215	2.86E-79	0.39701343	4.52E-75	Young	Cyst cells
RpS3	3.38E-23	0.39107893	5.34E-19	Young	Cyst cells
CycG	2.13E-09	0.44652577	3.36E-05	Young	Cyst cells
Pop2	2.90E-39	0.3314244	4.59E-35	Old	Cyst cells
RpII18	8.53E-107	1.49657794	1.35E-102	Young	Late spermatogonia
Rpb8	6.85E-97	1.33439304	1.08E-92	Young	Late spermatogonia
RpII15	8.27E-70	0.68675352	1.31E-65	Young	Late spermatogonia
Mes4	2.85E-46	0.26874512	4.51E-42	Young	Late spermatogonia
Rpb5	4.98E-37	0.30349123	7.88E-33	Young	Late spermatogonia
CycB	1.10E-08	0.38732343	0.00017418	Young	Late spermatogonia
timeout	1.21E-08	0.25715093	0.00019092	Young	Late spermatogonia
nej	1.94E-07	0.35125313	0.00307209	Young	Late spermatogonia
RpS3	2.40E-42	0.80385998	3.80E-38	Old	Late spermatogonia
14-3-3epsilon	7.92E-29	0.59657719	1.25E-24	Old	Late spermatogonia
Ubc6	1.12E-19	0.54643506	1.77E-15	Old	Late spermatogonia
Rad23	7.58E-16	0.49114339	1.20E-11	Old	Late spermatogonia
CSN8	2.26E-15	0.62861836	3.57E-11	Old	Late spermatogonia
Tctp	1.71E-13	0.33587259	2.70E-09	Old	Late spermatogonia

CG5316	8.06E-12	0.29793671	1.27E-07	Old	Late spermatogonia
Spt5	1.38E-11	0.29691068	2.19E-07	Old	Late spermatogonia
Kap-alpha3	2.63E-10	0.31897472	4.16E-06	Old	Late spermatogonia
Caf1-55	5.23E-10	0.48446497	8.28E-06	Old	Late spermatogonia
CG6171	7.90E-10	0.38454172	1.25E-05	Old	Late spermatogonia
RpLP0	1.30E-09	0.26187095	2.06E-05	Old	Late spermatogonia
CSN4	1.24E-08	0.41085392	0.00019591	Old	Late spermatogonia
Rcd5	3.17E-08	0.3099192	0.00050101	Old	Late spermatogonia
PCNA	6.53E-08	0.36666869	0.00103283	Old	Late spermatogonia
CG32756	1.02E-06	0.29439689	0.01615454	Old	Late spermatogonia
RpII18	0	2.24024073	0	Young	Late spermatocytes
Rpb8	0	0.68762069	0	Young	Late spermatocytes
RpS3	4.95E-196	1.37233373	7.83E-192	Old	Late spermatocytes
RpLP0	2.20E-180	1.25162021	3.47E-176	Old	Late spermatocytes
CkIalpha	7.80E-109	0.56529777	1.23E-104	Old	Late spermatocytes
CG42300	1.70E-103	0.50845639	2.68E-99	Old	Late spermatocytes
CG42299	8.67E-85	0.46817229	1.37E-80	Old	Late spermatocytes
SkpA	5.65E-81	0.33964967	8.93E-77	Old	Late spermatocytes
CSN6	4.23E-80	0.56697157	6.69E-76	Old	Late spermatocytes
CycB	1.28E-79	0.65484903	2.03E-75	Old	Late spermatocytes
Cdk1	4.13E-79	0.335287	6.53E-75	Old	Late spermatocytes
MED4	1.05E-78	0.33115862	1.66E-74	Old	Late spermatocytes
14-3-3epsilon	3.57E-78	0.55443737	5.65E-74	Old	Late spermatocytes
PCNA	1.47E-77	0.3804881	2.32E-73	Old	Late spermatocytes
Rad23	6.26E-71	0.66761187	9.91E-67	Old	Late spermatocytes
Nse4	1.43E-69	0.26158254	2.26E-65	Old	Late spermatocytes
CG5316	4.34E-66	0.26243411	6.87E-62	Old	Late spermatocytes
EndoGI	1.24E-65	0.36206536	1.97E-61	Old	Late spermatocytes
rept	5.50E-62	0.60888586	8.70E-58	Old	Late spermatocytes
Caf1-55	1.81E-58	0.26819856	2.86E-54	Old	Late spermatocytes
Xpc	1.10E-56	0.28523331	1.75E-52	Old	Late spermatocytes
CSN5	1.07E-55	0.31210701	1.69E-51	Old	Late spermatocytes
alien	1.81E-52	0.28582187	2.86E-48	Old	Late spermatocytes
lwr	2.13E-50	0.49797956	3.37E-46	Old	Late spermatocytes
vtd	3.44E-49	0.26127168	5.44E-45	Old	Late spermatocytes
E(bx)	3.98E-36	0.25696438	6.29E-32	Old	Late spermatocytes
CycG	6.96E-32	0.56181969	1.10E-27	Old	Late spermatocytes
Kap-alpha3	1.16E-29	0.39730812	1.83E-25	Old	Late spermatocytes
Ubc6	4.40E-28	0.3411957	6.95E-24	Old	Late spermatocytes

SMC5	1.50E-27	0.25141592	2.37E-23	Old	Late spermatocytes
pont	7.89E-26	0.37201391	1.25E-21	Old	Late spermatocytes
Not1	5.86E-15	0.30481652	9.27E-11	Old	Late spermatocytes
CSN7	8.00E-12	0.25165343	1.27E-07	Old	Late spermatocytes
His2Av	1.29E-11	0.26252299	2.03E-07	Old	Late spermatocytes
RpII18	1.06E-55	1.62277144	1.68E-51	Young	Hub cells
Rpb8	8.05E-21	0.75430337	1.27E-16	Young	Hub cells
RpLP0	3.40E-51	1.0391812	5.38E-47	Old	Hub cells
RpS3	8.87E-42	0.9375168	1.40E-37	Old	Hub cells
Rad23	2.10E-23	0.30312851	3.33E-19	Old	Hub cells
Ubc6	2.46E-23	0.27630497	3.89E-19	Old	Hub cells
Tctp	2.74E-10	0.38229026	4.34E-06	Old	Hub cells
14-3-3epsilon	3.66E-08	0.25003532	0.00057825	Old	Hub cells
RpII18	1.02E-29	1.8085739	1.61E-25	Young	Late spermatids
Rpb8	6.59E-10	0.5082588	1.04E-05	Young	Late spermatids
CG6171	1.17E-17	0.28246111	1.85E-13	Old	Late spermatids
Ubc6	4.51E-17	0.41362848	7.14E-13	Old	Late spermatids
CG42299	1.18E-16	0.30337013	1.86E-12	Old	Late spermatids
RpLP0	2.51E-15	0.53124104	3.98E-11	Old	Late spermatids
Not1	2.84E-15	0.50772477	4.49E-11	Old	Late spermatids
RpS3	1.07E-14	0.55119869	1.70E-10	Old	Late spermatids
jnj	3.92E-14	0.26369859	6.20E-10	Old	Late spermatids
CSN6	6.65E-14	0.29107235	1.05E-09	Old	Late spermatids
me31B	1.56E-10	0.30481526	2.46E-06	Old	Late spermatids
14-3-3epsilon	3.38E-10	0.51087335	5.35E-06	Old	Late spermatids
His2Av	9.01E-10	0.35886244	1.43E-05	Old	Late spermatids
corolla	1.17E-09	0.39476222	1.85E-05	Old	Late spermatids
CG10694	2.03E-08	0.33108618	0.00032075	Old	Late spermatids
RpII18	5.49E-70	1.22697945	8.68E-66	Young	Epithelial cells
Rpb8	1.80E-34	0.59744312	2.84E-30	Young	Epithelial cells
Ubc6	2.14E-11	0.43927645	3.39E-07	Old	Epithelial cells
CkIalpha	7.18E-07	0.26029988	0.01134822	Old	Epithelial cells
RpII18	4.22E-23	1.83948198	6.68E-19	Young	GSC, Early spermatogonia
Rpb8	3.76E-22	1.75808203	5.95E-18	Young	GSC, Early spermatogonia
Rpb5	1.25E-20	1.06269526	1.97E-16	Young	GSC, Early spermatogonia
RpII15	2.46E-20	1.22337098	3.89E-16	Young	GSC, Early spermatogonia
Mes4	8.30E-15	0.43020964	1.31E-10	Young	GSC, Early spermatogonia
DNApol-iota	1.15E-13	0.42905041	1.83E-09	Young	GSC, Early spermatogonia
E2f2	2.51E-13	0.91378164	3.97E-09	Young	GSC, Early spermatogonia

CG10694	7.27E-13	1.21320657	1.15E-08	Young	GSC, Early spermatogonia
Pol32	8.57E-13	0.33711043	1.36E-08	Young	GSC, Early spermatogonia
Nse4	4.15E-12	0.74620555	6.56E-08	Young	GSC, Early spermatogonia
CSN6	1.19E-11	0.81177098	1.88E-07	Young	GSC, Early spermatogonia
CG42300	1.42E-11	0.80520661	2.25E-07	Young	GSC, Early spermatogonia
RpII215	1.48E-11	0.70083696	2.34E-07	Young	GSC, Early spermatogonia
CG42299	2.96E-11	0.99282304	4.69E-07	Young	GSC, Early spermatogonia
CG3448	2.10E-10	0.3601618	3.32E-06	Young	GSC, Early spermatogonia
Rcd5	2.08E-09	0.63918995	3.29E-05	Young	GSC, Early spermatogonia
MED4	3.63E-09	0.58272682	5.74E-05	Young	GSC, Early spermatogonia
cerv	1.02E-08	0.41235174	0.00016179	Young	GSC, Early spermatogonia
Xpac	7.18E-08	0.5003643	0.00113534	Young	GSC, Early spermatogonia
Ubc6	7.57E-08	0.51802382	0.00119779	Young	GSC, Early spermatogonia
mu2	8.24E-08	0.28408084	0.00130277	Young	GSC, Early spermatogonia
phr6-4	1.04E-07	0.25882333	0.00163731	Young	GSC, Early spermatogonia
Topors	4.08E-07	0.37202306	0.00645265	Young	GSC, Early spermatogonia
Debcl	1.44E-06	0.35200272	0.02285286	Young	GSC, Early spermatogonia
RpS3	5.57E-10	1.13656114	8.81E-06	Old	GSC, Early spermatogonia
14-3-3epsilon	2.18E-08	0.67850579	0.00034418	Old	GSC, Early spermatogonia
Tctp	3.67E-07	0.58039379	0.00580205	Old	GSC, Early spermatogonia
RpLP0	2.34E-06	0.68157931	0.03708578	Old	GSC, Early spermatogonia

Supplemental table 1: Table of differentially expressed genome maintenance genes. These were calculated by splitting the main Seurat object into each cell type and setting the Idents to "age".

Gene	P value	Log_2 fold change	Bonferroni-adjusted P	Fly age	Cell type	Gene class
dmel-testis-AG-merged.15917	3.38E-49	0.29428559	5.33E-45	Young	Early spermatids	segregating
dmel-testis-AG-mergedplus.2408	5.55E-20	0.29227805	8.75E-16	Young	Early spermatocytes	segregating
dmel-testis-AG-mergedminus.14760	3.92E-09	0.3235566	6.18E-05	Young	Late spermatogonia	segregating
dmel-testis-AG-merged.15917	1.70E-16	0.26208888	2.67E-12	Old	Late spermatids	segregating
dmel-testis-AG-mergedplus.3086	6.28E-195	1.28564792	9.90E-191	Old	Early spermatids	fixed
dmel-testis-AG-merged.15917	1.17E-40	0.33935426	1.84E-36	Old	Early spermatids	fixed
CG43760	8.56E-87	0.5905714	1.35E-82	Young	Early spermatids	fixed
CG43750	3.19E-46	0.25678987	5.02E-42	Young	Early spermatids	fixed
CG43449	3.44E-41	0.2949748	5.42E-37	Young	Early spermatids	fixed
CG44174	2.47E-149	1.19382429	3.89E-145	Old	Early spermatocytes	fixed
CG44227	2.61E-27	0.29089427	4.12E-23	Old	Early spermatocytes	fixed
CG43760	1.16E-10	0.26595429	1.83E-06	Young	Early spermatocytes	fixed

CG43750	8.31E-154	0.93040937	1.31E-149	Old	Cyst cells	fixed
CG43449	3.10E-91	0.74957663	4.89E-87	Old	Cyst cells	fixed
CG43760	5.47E-47	0.96654881	8.62E-43	Old	Late spermatocytes	fixed
CG43750	1.01E-24	0.94625088	1.59E-20	Old	Epithelial cells	fixed
CG43760	1.27E-08	0.4154728	0.00019959	Old	Epithelial cells	fixed
CG43760	5.91E-70	1.07260707	9.31E-66	Old	Hub cells	fixed
CG43750	1.78E-25	0.44217757	2.81E-21	Old	Hub cells	fixed
CG43760	1.26E-48	0.99877606	1.98E-44	Old	Late spermatogonia	fixed
dmel-testis-AG-merged.2415	4.16E-17	0.44607375	6.55E-13	Old	Late spermatogonia	fixed
CG43760	5.80E-42	1.28883889	9.15E-38	Old	Late spermatids	fixed
CG44329	1.08E-15	0.29891384	1.71E-11	Old	Late spermatids	fixed
CG43760	6.07E-27	1.9003241	9.56E-23	Young	Late spermatids	fixed
dmel-testis-AG-merged.7007	1.94E-11	0.91881875	3.06E-07	Young	GSC, Early spermatogonia	fixed

Supplemental Table 2: Differentially expressed *de novo* genes. Unannotated *de novo* transcripts contain the string "dmel-testis-AG-merged" and correspond to entries in our custom reference gtf file using *de novo* transcripts published in our previous work (Witt et al., 2019).

Name	p_val	avg_log2FC	p_val_adj	cluster	celltype
DM412	4.45E-17	0.28765355	7.02E-13	Old	Early spermatids
Jockey2	5.23E-28	0.40396657	8.25E-24	Old	Early spermatocytes
Jockey2	2.63E-37	0.60405864	4.15E-33	Young	Early spermatocytes
BS2	4.30E-20	1.04723415	6.78E-16	Young	Early spermatocytes
DMRT1B	1.61E-79	0.50192081	2.53E-75	Old	Cyst cells
DOC	1.37E-62	0.26769752	2.16E-58	Old	Cyst cells
Jockey2	9.17E-62	0.26235339	1.45E-57	Old	Cyst cells
TAHRE	3.06E-61	0.28940335	4.82E-57	Old	Cyst cells
DMRT1B	4.01E-53	0.2895345	6.32E-49	Old	Cyst cells
HETA	2.91E-52	0.34062078	4.59E-48	Old	Cyst cells
TRANSIB2	2.63E-36	0.87271865	4.14E-32	Old	Late spermatocytes
DOC	3.39E-33	0.7125252	5.34E-29	Old	Late spermatocytes
DMRT1B	2.44E-24	0.54239771	3.84E-20	Old	Late spermatocytes
Jockey2	5.76E-11	0.29940896	9.07E-07	Old	Late spermatocytes
DM412	1.74E-07	0.29024674	0.00273574	Old	Late spermatocytes
TAHRE	2.69E-09	0.37897876	4.24E-05	Old	Epithelial cells
DOC	6.17E-32	0.53975512	9.73E-28	Old	Hub cells
Jockey2	1.91E-16	0.28563862	3.00E-12	Old	Hub cells
Jockey2	5.22E-08	1.12125594	0.00082305	Young	Late spermatogonia

DOC	3.03E-07	1.88372443	0.00477686	Young	Late spermatogonia
BS2	7.18E-19	0.3484718	1.13E-14	Old	Late spermatids
DMRT1B	4.46E-16	0.39895427	7.02E-12	Old	Late spermatids
DOC	1.33E-13	0.42069063	2.09E-09	Old	Late spermatids
DM412	1.53E-11	0.29732294	2.41E-07	Old	Late spermatids
Jockey2	3.01E-10	0.56404339	4.74E-06	Old	GSC, Early spermatogonia

Supplemental Table 3: Table of differentially expressed transposable elements.