UNDERSTANDING GENOME REGULATION OF THE AGING DROSOPHILA EYE

by

Juan Pablo Jauregui Lozano

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

West Lafayette, Indiana

May 2022

To my parents, Alicia and Enrique, and my sister, Natalia, who inspire me to be the best version of myself every day.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

ABSTRACT

Aging is characterized by changes in transcriptional outputs that correlate with physiological changes observed as we age, including decreased function, and increased cell death. Importantly, many of these changes are conserved across tissues and organisms . Because one of the molecular hallmarks of aging is epigenetic dysregulation, we are interested in understanding how age-associated changes in chromatin contribute to the aging transcriptome. To accomplish this, we use the *Drosophila* visual system as a model for aging, with a particular focus on photoreceptor neurons.

To perform cell-type specific genomic studies in *Drosophila,* we previously developed a nuclei immuno-enrichment method that was compatible with RNA-seq. However, due to low nuclei yields, this protocol was not amenable to chromatin-based studies, such as ChIP-seq and ATAC-seq. In Chapter 1, we developed an improved approach to isolate Drosophila melanogaster nuclei tagged with a GFP^{KASH} protein that increased yields without compromising efficiency. We further demonstrate that this protocol is compatible with several chromatin profiling techniques, such as Assay of Transposable-Accessible Chromatin (ATAC)-seq, Chromatin Immunoprecipitation (ChIP-seq), and CUT&Tag.

Chromatin accessibility is enriched for transcription factors. Thus, in Chapter 2, we profiled accessible chromatin in aging photoreceptors and integrated this data with RNA-seq to identify transcription factors that showed differential activity in aging Drosophila photoreceptors. Surprisingly, we found that 57 transcription factors had differential binding activity during aging, including two circadian regulators, Clock and Cycle, that showed sustained increased activity during aging. When we disrupted the Clock:Cycle complex by expressing a dominant negative version of Clock (ClkDN) in adult photoreceptors, we observed changes in expression of 15–20% of genes including key components of the phototransduction machinery and many eye-specific transcription factors. Using ATAC-seq, we showed that expression of ClkDN in photoreceptors leads to changes in activity of 37 transcription factors and causes a progressive decrease in global levels of chromatin accessibility in photoreceptors. Supporting a key role for Clock-dependent transcription in the eye, expression of ClkDN in photoreceptors also induced light-dependent

retinal degeneration and increased oxidative stress, independent of light exposure. Together, our data suggests that the circadian regulators Clock and Cycle act as neuroprotective factors in the aging eye by directing gene regulatory networks that maintain expression of the phototransduction machinery and counteract oxidative stress.

Previous work in the Weake lab found that long, highly expressed genes were more susceptible to be downregulated with age. DNA:RNA hybrids are co-transcriptional structures that form when the nascent RNA hybridizes with the template strand, resulting in a displaced non-template ssDNA. Importantly, accumulation of R-loops is associated with transcriptional inhibition and genomic instability, both hallmarks of aging. In Chapter 3, I characterized R-loop in maintaining proper transcriptional outputs and regulating visual function during aging. Bulk assays to measure R-loop levels revealed a significant increase in nuclear R-loops with age. Further, genome-wide mapping of R-loops revealed that transcribed genes accumulated R-loops over gene bodies during aging, which correlated with decreased expression of long and highly expressed genes. Importantly, while photoreceptor-specific down-regulation of Top3β, a DNA/RNA topoisomerase associated with R-loop resolution, lead to decreased visual function, over-expression of Top3β or nuclear-localized RNase H1, which resolves R-loops, enhanced positive light response during aging.

Together, these studies underscore the importance of understanding how age-related changes in genomic processes, such as circadian transcription and maintenance of R-loops, contribute to physiological changes observed during aging.

STATEMENT OF PUBLISHED WORK

Chapter 1 is a pre-copyedited manuscript published in Genetics in May, 2021, and available at https://doi.org/10.1093/genetics/iyab079. Chapter 2 is a pre-copyedited manuscript published in PLOS Genetics, and available at https://doi.org/10.1371/journal.pgen.1010021. Chapter 3 is a pre-copyedited manuscript published in Aging Cell, and available at https://doi.org/10.1111/acel.13554.

INTRODUCTION

Aging is defined as "a persistent decline in the age-specific fitness components of an organism due to internal physiological deterioration" (1). Importantly, studies across multiple organisms have shown that aging is accompanied by several molecular phenotypes, including dysregulation of protein homeostasis, metabolic imbalance and differential epigenetic regulation of gene expression (2). Defining precise molecular mechanisms that contribute to age-associated changes has been proven to be difficult, partially because majority of aging studies have been focused on whole organisms (3). Importantly, there is growing amount of evidence that supports a model of different tissues aging at a different rate (4). However, cell-type specific studies during aging are difficult to perform in vertebrate aging models and studies based on tissue dissection tend to buffer cell-type specific changes because of cellular heterogeneity and contamination with neighboring tissues (5), therefore making it nearly impossible to determine how each cell type responds and adapts to aging itself. Thus, these observations highlight the need to perform aging studies in the context of single tissues or cell types.

In our lab, we are interested in understanding the biology of the aging eye. One of the hallmarks of the aging eye, as well as many age-related eye diseases, is the loss of particular cell types within the retina, including photoreceptor neurons, such as rods and cones (6,7). While recent studies in mice have shown that vision can be restored by inducing re-differentiation of precursor cells into mature photoreceptors by expressing reprogramming transcription factor cocktails (8,9), we still have a primitive understanding of the genomic regulation processes that are required to maintain photoreceptor health during aging. Importantly, defining these processes may provide avenues to prevent or slow-down photoreceptor degeneration, leading to reduced incidence of age-related eye diseases.

Proper maintenance of transcriptional outputs is critical, because dysregulation of transcription can lead to altered cellular states, leading to disease. Further, proper transcriptional outputs require a tight regulation of every step, such including initiation, elongation, and termination. However, not much is known about how age affects the transcriptional machinery. We and others have shown that the transcriptional processes, such as splicing, become affected with age (2). Additionally, decreasing transcript levels of splicing factors, which lead to dysregulation of splicing outcomes,

in an eye-specific manner, lead to accelerated decrease in visual function, mimicking what is seen in older flies with normal levels of splicing factors (10). However, to our knowledge, there is no comprehensive understanding of how transcription factors, which initiate transcription by recruiting the RNA Polymerase II complex, contribute to the aging transcriptome. Thus, understanding what are the mechanisms that contribute to the maintenance of transcriptional outputs in visual neurons is an outstanding question in the field of biology of the eye (Figure 1). Since we and others have shown that modulation of factors associated with transcriptional regulation can lead to changes in eye physiology, we are interested in gaining a better understanding of the molecular changes that contribute to age-associated loss of photoreceptor function and survival. To study this, we use the *Drosophila* eye as model, with a focus on outer photoreceptor neurons. Outer photoreceptor neurons are homologs to mammalian rods, which are the photoreceptors involved in mediating peripheral vision, and also more abundant relative to cone cells (11). In order to perform our transcriptional studies in a tissue-specific manner, we previously developed a tissue-specific nuclear immuno-precipitation (NIE) approach that was compatible with RNA-seq (12), and we have extensively used this approach to profile the transcriptome of photoreceptor neurons under aging and blue-light stress (13,14). Additionally, this approach has been adapted in other labs to profile the transcriptome of different cell types (15–18). However, our previous protocol yielded low nuclei numbers, which made performing chromatin profiling and obtaining material from rare cell populations challenging.

My thesis was divided in firstly improving our tissue-specific nuclei immuno-enrichment approach, and validate several approaches to profile genome-wide aspects of chromatin biology, such as chromatin accessibility and histone modifications. Then, I applied these methods to aging photoreceptor neurons, and discovered a role for circadian transcription factors in promoting visual health by preventing light-dependent retinal degeneration and contributing to the response to increasing oxidative stress, which is often a hallmark of aging. Last, in collaboration with Dr. Hana Hall, I uncovered a novel role for regulation of RNA:DNA hybrid, or R-loop metabolism in maintenance of visual function during aging.
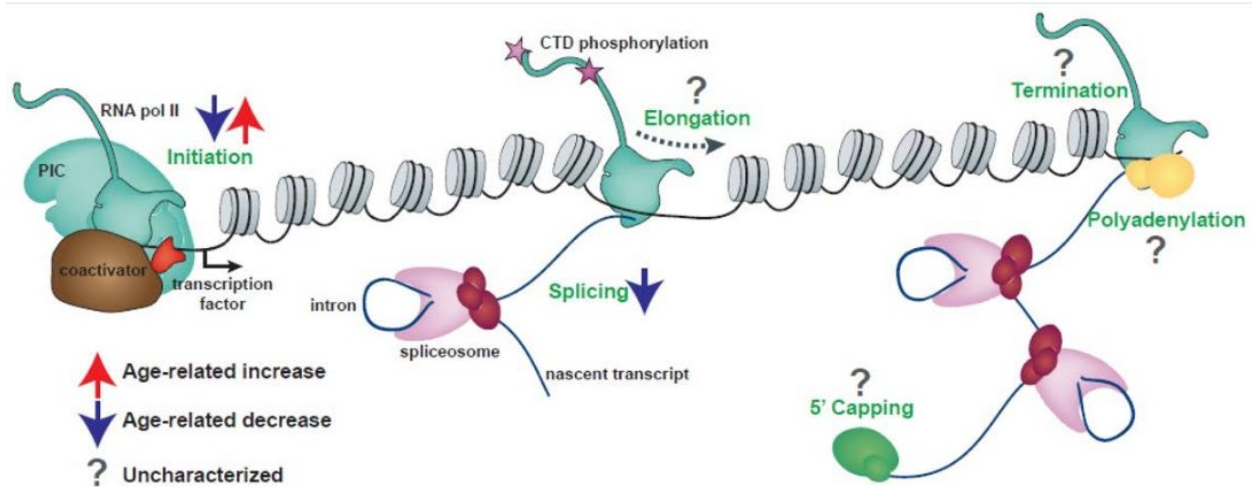
Figure 1. Age-associated changes to the transcriptional machinery (taken from (2)).

Age-associated impairment of transcriptional regulation processes might contribute to the aging transcriptome of Drosophila photoreceptor neurons. We and others have shown that 10-15% of the transcriptome has differential gene expression with age. In addition, we have observed dysregulation of splicing patterns. However, studies further evaluating age-associated changes in processes such as transcription factor occupancy, elongation rate, or termination efficiency, are lacking.

CHAPTER 1.   IN VIVO TISSUE-SPECIFIC CHROMATIN PROFILING IN DROSOPHILA MELANOGASTER USING GFP-TAGGED NUCLEI.

Declaration of collaborative work

The work described in this chapter was the collaborative work of Juan Jauregui-Lozano and Kimaya Bakhle, under the supervision of Vikki M. Weake. Kimaya Bakhle performed the quantitative PCR experiments. All other experiments, as well as data visualization were performed by Juan Jauregui-Lozano. The original draft was written by Juan Jauregui-Lozano, and edited by Juan Jauregui-Lozano and Vikki Weake

1.1   Introduction

Dynamic regulation of the epigenome is crucial to replication, transcription, and DNA repair. For instance, accessible chromatin is associated with gene regulatory sequences, such as enhancers, promoters and transcription factor binding sites, and contributes to transcription initiation [19]. In addition, chromatin-associated proteins, such as histones, transcription factors or chromatin remodelers, modulate several processes, including nucleosome occupancy [20], heterochromatin maintenance [21], and recruitment of DNA repair factors [22]. Thus, genome-wide chromatin profiling across different physiological states can help us understand how chromatin-mediated processes impact cell homeostasis.

The wide array of genetic manipulation tools, a highly mapped and annotated genome, relatively short lifespan, and ease of growth have made *Drosophila* one of the most widely used model organisms for studying the basic molecular mechanisms of eukaryotic cells [23]. Further, the tissue homology between *Drosophila* and humans can be leveraged to uncover regulatory mechanisms associated with human relevant conditions, such as aging, neurodegeneration, and diabetes [24–27]. Since epigenetic dysregulation is one of the hallmarks of many diseases, including cancer and neurodegeneration [28,29], profiling chromatin states in a tissue-specific context using *Drosophila* might improve our understanding of how chromatin-associated changes contribute to disease onset. However, profiling cell type-specific chromatin states *in vivo* is challenging. Although tissue dissection can be coupled with bulk and single-cell genome wide experiments, manual tissue dissection is technically demanding and contamination from surrounding tissues can often confound results.

To overcome these limitations, alternative techniques have been developed based around epitope labeling of nuclei by transgenic expression of an epitope tag driven by a cell-type specific promoter followed by purification (30). This approach has been coupled with Fluorescence-Activated Cell Sorting (FACS)-based nuclei isolation, such as the "Batch Isolation of Tissue-Specific" (BiTS) approach (31), as well as bead-based purification, such as the "Isolation of Nuclei Tagged in specific Cell Types" (INTACT) method (32). In addition, INTACT has been applied to tissue-specific experiments in *Arabidopsis* (33,34), *Drosophila* (35–38), *Xenopus* (39), and mice (40); FACS-based BiTS has also been extensively applied to purify different tissues from *Drosophila* embryos and mice (41). In *Drosophila*, these nuclei labeling approaches often rely on genetic tools for binary expression of transgenes, such as the well-established Gal4-UAS system (42). Currently, more than 8000 stocks that express Gal4 under control of different cell-type specific promoters are available through the Bloomington *Drosophila* Stock Center (BDSC). Thus, these nuclei tagging approaches combined with the Gal4-UAS expression system provide a powerful and flexible tool to manipulate and examine many cell-types in *Drosophila*.

We previously developed a Gal4-UAS based nuclei immuno-enrichment (NIE) protocol to isolate nuclei from specific *Drosophila* cell types labeled with an outer nuclear membrane localized GFP^{KASH} protein (12,13). The GFP^{KASH} protein consists of the Klarsicht, ANC-1, Syne Homology (KASH) domain of Muscle-specific protein 300 kDa (Msp300; FBgn0261836) fused to the C-terminus of EGFP, and localizes GFP to the outer nuclear membrane facing the cytoplasm (43). This GFP^{KASH}-based NIE approach was successfully applied to transcriptomic studies in specific cell populations, such as larval glial cells (44), adult photoreceptor neurons (13,14), and olfactory sensory neurons (15). However, our previous protocol yielded low nuclei numbers, which made performing chromatin profiling and obtaining material from rare cell populations challenging. In this study, we sought to optimize the NIE protocol to increase nuclei yield and stringency over background. Using this 'improved' GFP^{KASH}-based NIE protocol, we applied transcriptome and chromatin profiling techniques to NIE-purified adult *Drosophila* photoreceptor nuclei. We demonstrate the reproducibility and quality of datasets obtained profiling nuclear RNA-seq, an improved Assay for Transposable Accessible Chromatin followed by sequencing (ATAC-seq), called Omni-ATAC, and Chromatin immunoprecipitation followed by sequencing (ChIP-seq) and Cleavage Under Targets and Tagmentation (CUT&Tag) of histone modifications.

1.2   Results

1.2.1   Optimization of tissue-specific nuclei immuno-enrichment (NIE) from adult *Drosophila*

As a starting point for profiling chromatin states in specific cell types in *Drosophila*, we sought to improve nuclei yields obtained with the NIE protocol using flies that express the GFP^{KASH} tag in outer photoreceptor neurons driven by Rh1-Gal4 (herein referred as Rh1>GFP^{KASH}) (Mollereau et al., 2000). We reasoned that isolating nuclei in a buffer designed to retain the integrity of the nuclear envelope would increase the availability of the GFP^{KASH} epitope, which is anchored to the outer nuclear membrane with GFP facing the cytoplasm (46). Previous studies have shown that perinuclear proteins are retained when nuclei are purified using a detergent-containing isotonic buffer (47), suggesting that the outer nuclear membrane remains intact under these conditions. Based on this rationale, we replaced the hypotonic/hypertonic buffers used in the homogenization, incubation, and washing steps of our previous NIE method with detergent-containing isotonic buffers. We also decreased the relatively high concentration of NP-40 detergent used for homogenization during the immunoprecipitation steps to decrease background binding (see methods). We refer to our previous and new NIE approaches as the 'standard' and 'improved' methods, respectively (Figure 1A).

We first assessed how nuclei yields varied based on the NIE method used. To do this, we quantified DNA after each NIE reaction (n=4). We used DNA yield as a measure of nuclei yield because the magnetic beads used in the NIE auto fluoresce, making it difficult to quantify nuclei accurately using microscopy-based techniques (Figure 1B). The 'improved' method yielded 1.2 ng of DNA per fly, compared to 0.2 ng of DNA for the 'standard' method (Figure 1C). Considering that there are ~7200 outer photoreceptors per fly, and that a diploid *Drosophila* nucleus typically contains ~0.36 pg DNA (48), the 'improved' method yields around 45% of the tagged nuclei compared with 13% for the 'standard' approach. We note that the starting material for each NIE reaction was 400 age-matched Rh1>GFP^{KASH} flies homozygous for both Gal4 and UAS transgenes; nuclei yield decreased approximately two-fold when GFP^{KASH}-based NIE was performed using flies heterozygous for both transgenes, suggesting that higher GFP^{KASH} expression levels can further improve purification efficiency.

Next, we evaluated if the NIE-purified nuclei were enriched relative to background cell types. To do this, we mixed an equivalent number of Rh1>GFP^{KASH} flies with Rh1>mCherry-FLAG^{KASH}, performed GFP-based NIE, and extracted DNA before (PRE) and after (POST) immuno-enrichment. We then quantified the relative genomic copies of GFP and mCherry in each sample using quantitative PCR (qPCR). If nuclei from the POST sample are depleted of the mCherry^{KASH}-positive nuclei upon GFP-based NIE, then the ratio of GFP/mCherry for the POST sample will be higher than the value of one observed in the PRE sample, which contains an equivalent number of GFP and mCherry labeled nuclei. Using this approach, we observed 24-fold enrichment of GFP nuclei over mCherry using the 'improved' method, which compared favorably with the 20-fold enrichment observed using the 'standard' method (Figure 1D).

1.2.2 The transcriptome of nuclei purified with the "improved" approach is depleted of genes enriched in other cell types relative to the 'standard' approach

Because we had previously generated high-quality nuclear RNA-seq from outer photoreceptor nuclei using the 'standard' approach (13), we profiled the nuclear transcriptome of NIE-purified outer photoreceptor nuclei (Rh1>GFP^{KASH}) using the 'improved' method and compared the transcriptome between methods; we note that the identical genotype, sex, and age were used for both studies, and that both library sets were generated using the same amount of RNA. We first analyzed similarity between the two datasets by calculating Spearman correlation for gene counts (Figure 2A). Spearman's rank scores between replicates were high for both methods ($p<0.97$), and samples clustered together based on the method used for NIE. Further, we also observed similar clustering by NIE approach using Principal Component Analysis (PCA). Notably, the variation between biological replicates slightly decreased using the improved method.

The observation that samples clustered by method suggested there were differences between the datasets obtained using the different NIE methods. We sought to identify the differences in gene expression associated with each NIE method by analyzing differentially expressed genes (DEGs) (n=3). Surprisingly, we identified 2046 DEGs (FDR < 0.01, FC > 2) between the two NIE methods, despite their identical genotypes, sex, and age (Figure 2B). Amongst these genes, 824 genes were upregulated in the improved dataset, and 1224 genes were upregulated in the standard dataset, representing improved- or standard-enriched genes, respectively. RNA-seq libraries for each

experiment were made using different RNA-seq kits (see methods). Since we used a kit designed for low-input material (200 pg – 10 ng RNA) to make the improved dataset libraries, we wondered if genes enriched in the improved dataset were being quantified as lowly-expressed in the standard dataset. However, the identified DEGs spanned a wide range of expression levels, including low, medium, and highly-expressed genes (Figure 2C), suggesting that differences in amplification of lowly abundant transcripts do not account for the differences in expression observed between the two approaches. Instead, inspection of the top DEGs in each condition revealed that several rhodopsin genes (*Rh3, Rh4,* and *Rh6*) were enriched in the standard method relative to the improved method. These rhodopsin proteins are highly enriched in inner photoreceptors (R7-R8) and are also expressed in the Johnston organ (49,50), but are not expressed in outer photoreceptors; conversely, Rh1-Gal4 is expressed only in the outer photoreceptors (45). Since inner photoreceptor-specific genes were enriched in the standard dataset, these observations suggest that the 'improved' method yields a more tissue-specific enriched nuclei pool relative to our previous approach. GO-term analysis of genes that were upregulated in each dataset revealed that the standard-enriched DEGs were enriched for categories such as neuropeptide signaling pathway, muscle contraction, and muscle structure development (Figure 2D, top). Further, gene-concept network analysis revealed enrichment of 42 genes associated with non-photoreceptor cell types in the standard-enriched DEGs, including ventral lateral neuron-expressed *Pdf* (FBgn0023178), protocerebrum-enriched *Dsk* (FBgn0000500), and muscle-enriched *Unc-89* (FBgn0053519) (Figure 2E) (51,52). In contrast, GO terms over-represented in the improved-enriched DEGs were associated with processes related to protein folding and polytene chromosome puffing (Figure 2D, bottom). Gene-concept network analysis revealed that the over-representation of these GO-term categories were driven by a modest enrichment of five Heat Shock Protein (Hsp) genes (Figure 2F).

Because of a statistically significant over-representation of biological processes associated with other cell types in 'standard' enriched dataset, these findings suggest that nuclei purified using the 'improved' NIE method have higher enrichment of tissue-specific transcripts compared to the 'standard' approach. However, we note that these experiments were performed at different times, and therefore batch effects and other variables might also influence our observations. Considering that the 'improved' method also had higher nuclei yields, we proceeded to optimize the subsequent

chromatin profiling methods with NIE-purified outer photoreceptor nuclei from Rh1>GFP^{KASH} flies using this method.

1.2.3   Profiling chromatin accessibility (Omni-ATAC) in NIE-purified nuclei

We next sought to profile accessible chromatin of NIE-purified nuclei using Omni-ATAC, a recently modified ATAC-seq technique which yields higher quality data, especially with lower input (53). ATAC-seq techniques, including Omni-ATAC, require optimization of the number of nuclei or cells used for each reaction to generate appropriate DNA fragment sizes and avoid either under- or over-tagmentation. Normally, cultured cells are counted to achieve precise numbers of cells per assay. However, nuclei bound to magnetic beads cannot be easily quantified using an automatic cell counter because the free magnetic beads interfere with the identification of individual nuclei (see Figure 1B). To overcome this limitation, we isolated genomic DNA from a fraction of the purified nuclei and normalized input material for Omni-ATAC reactions based on this quantification (Figure 3A). We note that because our protocol begins with NIE-purified nuclei, mitochondria are already depleted from the initial starting material, as shown by qPCR analysis of mitochondrial DNA present in the PRE and POST NIE samples. To evaluate whether differences in starting material would substantially alter data quality, we performed Omni-ATAC using either 50 or 100 ng of DNA (corresponding to approximately 125,000 and 250,000 nuclei, respectively) with a fixed amount of Tn5.

Tapestation analysis of Omni-ATAC libraries revealed similar DNA laddering patterns with both amounts of input nuclei. We then sequenced these libraries, and evaluated the size distribution of the mapped fragments. We observed the expected nucleosomal phasing distribution in both libraries (Figure 3B), with the first peak (80-120 bp) corresponding to nucleosome-depleted region (NDR)-associated DNA, followed by a peak around 180 bp corresponding to mononucleosome-associated fragments. Genome browser inspection of the data revealed discrete peaks with similar enrichment profiles obtained under each condition (Figure 3C). Since the Omni-ATAC signal should be enriched around transcriptional start sites (TSS), we next evaluated read distribution around the TSS of protein-coding genes (Figure 3D). We observed a significant enrichment of Omni-ATAC signal around the TSS with no differences between the 50 ng- and 100 ng- associated

datasets. This finding was further corroborated by heatmap plots of all protein-coding genes ranked based on their Omni-ATAC signal enrichment around the TSS.

Next, we evaluated the genomic distribution of peaks from both samples (Figure 3E). As expected from the observed enrichment of Omni-ATAC signal around the TSS (Figure 3C), 70% of the peaks mapped to promoters with no discernible differences in distribution between the two samples, and samples had a Spearman's score of 0.98. Because accessible chromatin is enriched for active promoters, we next evaluated if chromatin accessibility levels correlated with transcript levels detected by nuclear RNA-seq (see Figure 2). To do this, we divided the 13930 genes in the *Drosophila* genome based on their position on the heatmap into six groups, where genes are ranked based on the Omni-ATAC signal around the TSS (Figure 3G), and plotted the transcript level ($\log_2$ transcript per million - TPM) for all genes in each cluster (Figure 3H). We observed a positive correlation between the levels of chromatin accessibility at the TSS and transcript expression levels. Altogether, these observations suggest that high-quality chromatin accessibility data can be obtained from NIE-purified nuclei using as little as 50 ng of DNA equivalent of starting material, when coupled with Omni-ATAC.

1.2.4 Omni-ATAC of NIE-purified nuclei does not require high sequencing depth

To benchmark the quality and reproducibility of the Omni-ATAC protocol using the NIE-purified nuclei, we sought to systematically evaluate different quality control metrics of ATAC-seq datasets. We performed Omni-ATAC on NIE-purified nuclei equivalent to 100 ng of DNA in four independent biological samples, processing and analyzing each replicate individually (n=4). We first calculated the Spearman's correlation based on read distribution over a 500-bp binned genome, and found high reproducibility between samples, with Spearman's p scores above 0.90 (Figure 4A). Next, we plotted the Omni-ATAC signal around the TSS of protein coding genes (Figure 4B). We observed that the enrichment profiles around the TSS were highly consistent between replicates, corroborating the Spearman's correlation analysis. Next, we sought to evaluate the quality of peak-based analysis for each sample. Genome browser inspection of Omni-ATAC signal next to the peaks corresponding to each replicate showed high consistency, as determined by signal intensity of peaks (Figure 4C). Further, 88% of peaks presented significant overlap amongst all four replicates (Figure 4C). Similarly, we observed high concordance by

Irreproducible Discovery Rate (IDR) analysis of peaks between replicates, with all pair-wise comparisons having an IDR value above 0.61. The Fraction of Reads in Peaks (FRiP) score is a common quality control metric for genomic datasets, such as ChIP-seq and ATAC-seq, that measures overall signal-to-background ratio, as defined by ENCODE guidelines (54). According to ENCODE, good quality ATAC-seq datasets are defined as having FRiP score higher than 0.3. Thus, we next evaluated how FRiP scores varied based on sequencing depth. To do this, we down-sampled each replicate to 0.5, 1, 2.5, 5, 10, 20, 30, 40, and 50 million mapped fragments, and obtained its corresponding FRiP score (Figure 4E). FRiP scores did not vary significantly between replicates, and surprisingly, there was no substantial improvement in FRiP scores past 10 million mapped fragments. Further, visual inspection of the down-sampled data on a genome browser revealed similar enrichment of peaks at only 0.5 million fragments, resembling that observed using 50 million fragments. Next, we evaluated the number of peaks called for each sample based on the number of fragments (Figure 4G). As expected, peak calling benefited from the higher sequencing depth. However, when the number of peaks identified was normalized to the sample with greatest sequencing depth (50 million mapped fragments), we found that obtaining 20 million fragments identified approximately 80% of all possible peaks. Taken together, these observations imply that Omni-ATAC datasets do not require high sequencing depth for consistent gene- and peak-based analysis, and that 10-20 million reads is likely sufficient for most peak-based analyses in *Drosophila* samples.

1.2.5   The histone methylation landscape of adult *Drosophila* photoreceptors

Chromatin Immunoprecipitation (ChIP) is one of the most commonly used techniques in the genomics field, whereby sonicated chromatin is used to immunopurify a protein-DNA complex, followed by purification of the enriched DNA. Coupled with qPCR or high-throughput sequencing (ChIP-seq), it allows researchers to interrogate if a protein of interest is bound to a particular locus, or assay its genome-wide distribution, respectively. We sought to optimize a ChIP protocol suitable for use with NIE-purified nuclei. *During development of the protocol, w*e initially found that fixing the nuclei during homogenization led to an increase in background nuclei upon NIE (data not shown), leading us to modify the protocol so that the chromatin was cross-linked while the nuclei were immobilized on the magnetic beads, immediately following NIE (Figure 5A). Chromatin was then sonicated, and ChIP performed using standard approaches (see methods).

To benchmark the ChIP protocol, we examined genome-wide distribution of two histone methyl marks, Histone H3 Lysine 4 tri-methylation (H3K4me3) and H3 Lysine 36 tri-methylation (H3K36me3), both of which have been widely characterized by ChIP-qPCR and ChIP-seq studies in Drosophila and other organisms. We also examined the distribution of bulk histone H3, as well as an input sonicated chromatin control. First, we assessed the enrichment of each antibody by evaluating the overall distribution of reads over gene bodies for all protein-coding genes. Histone H3 is distributed throughout both active and repressed chromatin, and is usually slightly depleted around the TSS of transcribed genes (55). In *Drosophila,* as well as *in Saccharomyces cerevisiae* and in humans, H3K4me3 is enriched at the TSS whereas H3K36me3 localizes to gene bodies (56). Consistent with this expected distribution, we observed depletion of histone H3 and enrichment of H3K4me3 around the TSS, while H3K36me3 was enriched towards the 3′ region of the gene body (Figure 5B). Further, genome browser inspection of individual genes, such as the photoreceptor-enriched genes *trp* and *trpl,* corroborated the enrichment for H3K4me3 around the TSS and H3K36me3 over the gene body. In contrast, the inner photoreceptor-expressed *Rh3* showed no enrichment for either histone mark, as expected based on its lack of expression in outer photoreceptors (Figure 5C).

Next, we assessed the reproducibility between the replicates obtained using our ChIP-seq approach. Given the semi-quantitative nature of ChIP-seq, there has been growing interest in adding exogenous chromatin prior to immunoprecipitation, using the reads that map to the "reference" genome for spike-in normalization (57). To facilitate this spike-in normalization approach, we added 5% of *Arabidopsis thaliana* chromatin to *Drosophila* samples before each immunoprecipitation. To evaluate how the similarity between individual samples varied based on the normalization method, we normalized the data using the *Arabidopsis* spike-ins (as described in (58) or calculated traditional counts per million or CPMs. We then calculated the Spearman correlation of read coverage over the binned genome for H3K4me3 and H3K36me3 separately (Figure 5D-E). Interestingly, the H3K4me3 samples clustered based on the normalization method used, although there were no major differences between Spearman's rank scores obtained for individual samples using either approach. Replicate correlation was high for both normalization methods ($p > 0.96$ for both normalization methods). Strikingly, the H3K36me3 samples clustered together based on replicate rather than normalization approach, and each replicate had a *p=1,* with its normalization counterpart. Corroborating the heatmap findings, metaplot analysis of the

H3K4me3 distribution around the TSS and H3K36me3 distribution over gene bodies showed no substantial differences between biological replicates using either normalization approach. To further assess similarity between the replicates based on antibodies used, we next evaluated Spearman's correlation of CPM-normalized data for H3, H3K4me3, and H3K36me3 (Figure 5F). Corroborating the findings from the global read distribution over gene bodies, samples clustered together based on antibody. Moreover, the correlation between replicates for each antibody was also high ($p > 0.96$). We also found that normalizing histone methylation ChIP-seq data to the corresponding input did not affect the correlation findings based on samples clustering with antibodies, although Spearman's rank scores slightly decreased between biological replicates ($p > 0.71$, $p > 0.80$ and $p > 0.90$ for H3, H3K4me3 and H3K36me3, respectively). Moreover, genome browser inspection revealed that the H3 samples also presented the same amplification bias as input, suggesting that H3 normalization was sufficient for between-sample comparisons. Moreover, normalizing the histone methyl marks to histone H3 also controls for nucleosome occupancy.

Because H3K4me3 and H3K36me3 are histone modifications associated with active transcription, we next asked if H3K4me3 and H3K36me3 ChIP-seq signal levels positively correlated with gene expression. To do this, we ranked all protein-coding genes based on H3-normalized H3K4me3 signal around the TSS (Figure 5G, left) or H3-normalized H3K36me3 signal over gene bodies (Figure 5H, right), and separated all 13930 genes into six clusters based on their level of the respective histone mark. We then examined gene expression for each cluster by plotting transcript levels for each gene in the cluster (log_2 transcript per million -TPM) (Figure 5H). Similar to our observations for the Omni-ATAC clusters, H3K4me3 and H3K36me3 levels positively correlated with active transcription.

Overall, these observations demonstrate that chromatin obtained from NIE-purified nuclei accurately reflect the transcriptional state of these cells and can be used for profiling of chromatin accessibility and histone modifications. Furthermore, in our hands, adding a reference genome for spike-in normalization does not outperform traditional CPM normalization. We note that although the ChIP-seq data shown here was generated from libraries that used 2 ng of DNA as starting material, libraries made with as little as 100 pg of DNA showed comparable profiles, suggesting that this ChIP-seq protocol is amenable to low-input starting material. We also performed qPCR

on ChIP samples obtained using this protocol, demonstrating that this approach may be useful for researchers interested in examining individual genes rather than performing genome-wide studies.

1.2.6   NIE-purified nuclei are compatible with CUT&Tag for profiling histone marks

Last, we sought to apply CUT&Tag to NIE-purified nuclei. CUT&Tag is a recently developed technique used to profile chromatin, whereby a fusion protein (pAG-Tn5) targets an antibody-bound chromatin target, followed by tagmentation and release of enriched DNA (59). CUT&Tag has several advantages over ChIP-seq, including shorter sample processing times and lower background signal, therefore requiring less sequencing depth to identify high probability binding sites for proteins of interest. Further, CUT&Tag yields sequencing-ready libraries with no need for an additional library construction step. Based on these advantages, we sought to develop a CUT&Tag approach suitable for use with NIE-purified nuclei using commercially available Protein A/Protein G-Tn5 (pAG-Tn5).

Standard CUT&Tag protocols require cell/nuclei immobilization with Concanavalin A beads. However, NIE-purified nuclei are already bound to Protein G-magnetic Beads (PGBe), providing an initial starting point for CUT&Tag protocols. Our first H3K4me3 CUT&Tag trials with NIE-purified nuclei using PGBe were unsuccessful, and we wondered if the rabbit anti-H3K4me3 antibodies were being adsorbed by the excess protein G in our nuclei preparations (Figure 6A). To test this possibility, we performed NIE using Mouse IgG-coupled magnetic Beads (MIBe) instead of PGBe. Strikingly, performing NIE with MIBe led to successful purification of DNA following CUT&Tag, suggesting that PGBe were interfering with CUT&Tag steps. We then performed H3K4me3 CUT&Tag using age and sex-matched photoreceptor nuclei in order to compare the data with H3K4me3 ChIP-seq, since both datasets were obtained using the same antibody. TapeStation profiles of the four replicates detected sub-, mono- and di-nucleosomal fragments, with significant enrichment for mononucleosome-associated DNA (Figure 6B). We then proceeded with paired-end sequencing of the libraries. Genome browser inspection of CPM-normalized H3K4me3 CUT&Tag, H3K4me3 ChIP-seq, and Omni-ATAC signal (Figure 6C) revealed that H3K4me3 CUT&Tag signal had a more similar distribution to ChIP-seq than Omni-ATAC based on the number of "pseudo-peaks", or high signal points. Notably, background levels in H3K4me3 CUT&Tag were much lower than ChIP-seq, as shown by signal in intergenic regions.

To further assess the correlation between each CUT&Tag replicate, we calculated Spearman's correlation rank scores. Because CUT&Tag data had very low levels of background relative to ChIP-seq, we calculated the correlation based on read coverage over the narrow peaks obtained from the H3K4me3 ChIP-seq data instead of the binned genome. As expected from the above comparisons, samples clustered together based on technique. Using this approach, ChIP-seq samples had higher correlation values between individual replicates ($p>0.9$) compared with CUT&Tag replicates ($p>0.83$).

Since CUT&Tag replicates had a relative lower similarity compared to ChIP-seq, we next evaluated genome-wide signal distribution of H3K4me3 ChIP-seq and CUT&Tag around the TSS using metaplots. Signal distribution of H3K4me3 ChIP-seq around the TSS was highly consistent between replicates (Figure 6E-top), which further corroborated the Spearman's analysis. Surprisingly, H3K4me3 CUT&Tag signal was highly variable between replicates, both in intensity and signal distribution (Figure 6E-bottom). In addition, metaplots revealed that CUT&Tag replicates 1, 2 and 3 had a substantial increase in signal around the TSS that resembled the Omni-ATAC signal. To further assess to what extent the CUT&Tag signal originated from H3K4me3-associated DNA, we obtained heatmaps of CPM-normalized signal over gene bodies and compared ChIP-seq, CUT&Tag and Omni-ATAC. Heatmaps revealed that overall, CUT&Tag replicates showed similar signal distribution over gene bodies, with similar distribution to ChIP-seq (Figure 6F). Notably, Omni-ATAC signal was highly enriched around the TSS, with no significant enrichment over gene bodies. Last, we evaluated the signal to background ratio for H3K4me3 CUT&Tag data compared to H3K4me3 ChIP-seq using the CUT&Tag replicate that most closely resembled ChIP-seq (R4). To do this, we obtained FRiP scores for the H3K4me3 CUT&Tag replicate 4 and H3K4me3 ChIP-seq by down-sampling the H3K4me3 CUT&Tag-R4 and ChIP-seq samples to 0.5, 1, 2.5, 5, 10, and 15 million mapped fragments (Figure 6D). Notably, CUT&Tag substantially outperformed ChIP-seq with a FRiP score of 0.367 for CUT&Tag data even at only 0.5 million mapped fragments. In comparison, the FRiP score for ChIP-seq data only reached 0.266 at 15 million fragments.

Taken together, these observations indicate that a slight modification to the NIE reagents makes it possible to apply CUT&Tag to NIE-purified nuclei, providing a cost effective and efficient way of examining the genome-wide distribution of DNA-binding proteins. However, we note that the

increased variability observed between CUT&Tag replicates relative to ChIP-seq samples, as well as the presence of accessible DNA in some of the CUT&Tag datasets, suggests that further optimization to the protocol is required to improve the quality of the datasets.

1.3   Discussion

Here, we demonstrate the feasibility of chromatin profiling in specific cell types using immuno-enriched nuclei as starting material and show that profiling of chromatin accessibility and histone modifications associated with active transcription correlate with the transcriptional state of the profiled cell type. Our NIE approach enables isolation of nuclei within one hour, that can be subsequently used for RNA, DNA, and chromatin extraction, therefore enabling the application of RNA-seq, ATAC-seq, ChIP-seq, and CUT&Tag (Figure 7A). Compared to our 'standard' approach, the 'improved' NIE approach increased the nuclei yield 4-fold, as shown by genomic DNA quantification. In our hands, we have found that quantifying genomic DNA was the simplest and most robust approach to normalizing input amounts for samples because we could easily extract DNA from a small fraction of each sample, and process and quantify multiple samples within ~20 – 30 minutes. This allowed for robust and reproducible quantification of several samples at the same time. Nonetheless, it is possible to use different quantification techniques to assess nuclei yields, such as nuclei counting using a hemocytometer, since nuclei can be easily stained with DAPI and counted manually using a fluorescent microscope.

By isolating nuclei, rather than cells, we can obtain highly pure nuclear RNA that provides a view of the actively transcribed genome. While these data correlate with the adult photoreceptor transcriptome determined in our previous studies using a similar approach (13), our modified NIE protocol results in significant decrease in transcript levels of genes associated with other cell types. However, we cannot conclude on whether these differences are biological or technical, or both, because these experiments were performed at different times and there could be artifacts induced by differences in the experimental conditions that could also influence gene quantification, such batch effects and different library preparation kits. Combining this improved NIE approach with library construction kits developed for low RNA inputs, such as the one used in this study, will facilitate RNA-seq studies on much rarer cell populations, or on cells labeled in mosaic animals, that have previously been difficult to analyze using other techniques.

In addition to RNA-seq, we profiled accessible chromatin at a genome-wide scale in the NIE-purified nuclei using Omni-ATAC. To our knowledge, this is the first report of cell-type specific chromatin accessibility data in adult *Drosophila*, although ATAC-seq studies have been performed in different embryonic cell-types isolated using the INTACT method (36) and in dissected larval imaginal discs (60). Here, we show that using as little as 50 ng DNA equivalent of NIE-purified nuclei was sufficient to produce high-quality genome-wide chromatin accessibility data, suggesting that this technique should be suitable for lowly abundant cell types. Published reports have shown that ATAC-seq and Omni-ATAC can be applied to as little as 500 human cells (53,61), indicating that these chromatin profiling approaches are highly amenable to low input samples.

We also applied two approaches to profile genome-wide distribution of histone modifications, ChIP-seq and CUT&Tag. Our ChIP-seq protocol is amenable to incorporation of exogenous chromatin for spike-in normalization, although in our hands, normalizing the ChIP-seq data with a published spike-in normalization approach did not outperform traditional CPM normalization. We note that there has been discussion of the caveats for spike-in normalization with regard to ChIP-seq data (Refer to Dickson, 2020). Last, we switched the beads used for NIE from protein-G Dynabeads to mouse IgG Dynabeads, allowing successful application of H3K4me3 CUT&Tag to NIE-purified nuclei. To our knowledge, this work is the first report of tissue-specific CUT&Tag in *Drosophila*. Although the H3K4me3 CUT&Tag data showed increased variability, and yielded a combination of accessible and H3K4me3 associated DNA, FRiP score evaluation showed that even at a low sequencing depth ($1x10^6$ mapped fragments), H3K4me3 CUT&Tag signal-to-background ratio outperformed the ChIP-seq data obtained using the same antibody. We expect NIE-purified nuclei to be compatible with CUT&RUN techniques using a similar approach to that described in this study, since both techniques are based on the same principle; CUT&RUN uses MNase to digest and release enriched DNA (62).

Together, our data demonstrate that combining the improved NIE protocol with commonly used chromatin profiling techniques provides a feasible approach to characterizing the transcriptome and epigenome of specific cell types in *Drosophila*. In our hands, purifying nuclei using homozygous *Rh1-Gal4, UAS-GFP^{KASH}* yield 1.2 ng and 0.67 ng of DNA and RNA, respectively, per fly. Based on these estimations, we have calculated how many nuclei would be required to perform all the experiments presented in this manuscript (Figure 7B). In addition, the NIE

approach will facilitate researchers interested in profiling different aspects of chromatin biology using a single biological sample because purified nuclei can be split for RNA, DNA and chromatin extraction. Nuclei resuspended in Trizol or a Trizol-derivative are stable and can be kept for later RNA extraction. In addition, in our hands, nuclei can be incubated on ice for 1 hour before tagmentation without loss of chromatin integrity. Based on the wealth of available Gal4 drivers for cell type-specific expression in *Drosophila*, the NIE approach described here provides a flexible and resourceful chromatin profiling toolkit for researchers to interrogate chromatin-associated processes in a tissue-specific context. Additionally, we have generated fly stocks expressing the GFP[KASH] tag under the Q binary expression system (63) as well as UAS lines that tag nuclei with either mCherry-FLAG, 6xmyc or mCherry-FLAG/GFP, to provide additional flexibility for studies in *Drosophila.* These stocks are available at Bloomington Drosophila Stock Center (Figure 7C).

1.4   Methods

1.4.1   Fly strains

Flies homozygous for Rh1>GFP[KASH] = *P{ry[+t7.2]=rh1-GAL4}3, ry[506], P {w[+mC] = UAS-GFP-Msp300KASH}attP2* or Rh1>mCherryKASH, *P{ry[+t7.2]=rh1-GAL4}3, ry[506], P{w[+mC] = UAS-Msp300KASH-mCherry-Flag}attP2}* (Hall et al., 2017) were raised in 12:12 h light:dark cycle at 25°C on standard fly food. Flies were maintained in population cages with a density of ~1000 flies/cage. Fresh food was switched every other day. For all the biological replicates, male flies were collected at 10 days post-eclosion at Zeitgeber time 6 (-/+ 1 hour).

1.4.2   Nuclei Immuno-Enrichment (NIE)

NIE was performed as described previously (Hall et al., 2017; Ma & Weake, 2014) with minor modifications to the buffers used through-out the protocol. Briefly, fly heads from 400 age-matched flies were collected by freezing flies in 5 cycles of flash-freezing and vortexing. Fly heads were collected using frozen sieves and transferred to a 1 mL Dounce homogenizer containing 1 volume of homogenization buffer (40 mM HEPES, pH 7.5, 120 mM KCl and 0.4% v/v NP-40). Flies were homogenized using 10 strokes with 'loose pestle' followed by 10 strokes with 'tight' pestle. Homogenized lysate was then filtered using 40 µm cell strainers (Corning, Tewksbury MA,

Catalog# 352340), and NP-40 was diluted to 0.1% final concentration by adding 3 volumes of Dilution buffer (40 mM HEPES, pH 7.5 and 120 mM KCl). Nuclei were immuno-enriched using 40 µL of Dynabeads Protein G (ThermoFisher, Waltham MA, Catalog #10004D) pre-coupled with 4 µg of mouse anti-GFP antibody (Sigma Aldrich, St. Louis MO, Catalog #11814460001) for RNA-seq, ChIP-seq and Omni-ATAC experiments. For CUT&Tag, nuclei were immunoenriched using 40 µL of Dynabeads Pan Mouse IgG (ThermoFisher. Catalog #11041) pre-coupled with 4 □g of mouse anti-GFP antibody (Sigma Aldrich, Catalog #11814460001). Beads and nuclei were incubated at 4°C for 30 min with constant rotation, followed by 3 x 5-min washes with homogenization buffer at 4°C.

1.4.3 Quantitative PCR

DNA was purified with Quick-DNA Microprep Plus Kit (Zymo Research, Irvine CA, Catalog #D4074) and qPCR was performed using Bullseye EvaGreen qPCR 2X master mix-ROX (Midsci, Valley Park, MO, Catalog #BEQPCR-R) following the manufacturer's instructions.

1.4.4 RNA-seq

Purified nuclei were resuspended in 100 µL TRI reagent (Zymo Research, Catalog #R2050-1-200). RNA was purified using Direct-zol™ RNA Microprep (Zymo Research, Catalog, #R2061) and quantified with Qubit™ RNA HS Assay Kit. 10 ng of nuclear RNA were used for construction of cDNA libraries with Ovation SoLo RNA-seq System with Drosophila-specific anyDeplete technology for rRNA depletion (Tecan, Redwood City, CA, Catalog #0502-32). Up to 16 libraries were pooled in one lane for paired-end 150 bp Illumina HiSeq sequencing.

1.4.5 Omni-ATAC

Transposition was performed as published (Corces et al., 2017). Briefly, a fraction of immunoprecipitated nuclei were purified with Quick-DNA Microprep Plus kit (Zymo Research, Catalog #D4074). Nuclei corresponding to 50 or 100 ng were aliquoted and resuspended in 50 µL of Transposition mix (25 µL 2x TD buffer, 16.5 µL PBS, 0.05 µL 1% v/v Digitonin, 0.05 µL 10% v/v Tween and 2.5 µL TDE1 enzyme (Illumina, San Diego CA, Catalog #20034198). Tagmented DNA was purified with Zymo DNA clean & concentrator-5 kit (Zymo Research #D4013).

Libraries were constructed using IDT for Illumina Nextera DNA Unique Dual Indexes Set A (Illumina, Catalog #20027213) and 7 PCR cycles were used to amplified libraries using NEBnext High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich MA, Catalog #M0541S) and SYBR Green I (ThermoFisher, Catalog #S7563). To determine additional cycles, Nextera primers 1 and 2 were used. Purified libraries were submitted to a round of double-size selection with AMPure XP beads (Beckman Coulter, Brea CA, Catalog #A63880) with a 0.5X-1.0X ratio. The fragment size distribution of libraries was assessed with TapeStation High-Sensitivity D1000 Screentapes (Agilent, Santa Clara CA, Catalog #5067-5584). Up to 16 libraries were pooled in one lane for paired-end 150 bp Illumina HiSeq sequencing.

1.4.6 ChIP-seq

Chromatin extraction (Drosophila): Immunoenriched nuclei were resuspended in 1 mL of A1 buffer (15 mM HEPES, pH 7.5, 15 mM NaCl, 60 mM KCl, 4 mM $MgCl_2$, 0.5% Triton X-100 v/v) and cross-linked with 1% methanol-free formaldehyde (ThermoFisher #28906) for 2 min at room temperature. Fixed nuclei were quenched with 125 mM Glycine, pH 7.5 for 5 min, followed by sonication in 130 µL of Nuclei Lysis Buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% v/v SDS) in Covaris E220 with the following conditions: 10 min, 2% duty cycle, 105 Watts and 200 c.p.b. to obtain an average fragment size of ~320 bp. Chromatin was centrifuged at 14,000 rpm, 10 min, 4°C, and the soluble chromatin supernatant was diluted with X-ChIP dilution buffer (16.7 mM Tris, pH 8.0, 167 mM NaCl, 1% Triton X-100 v/v, 1.2 mM EDTA pH 8.0), flash-frozen in liquid nitrogen, and stored at -20°C. Chromatin extraction (Arabidopsis): 2.5 g of 10-day old *ref4-3MED15FLAG Arabidopsis thaliana* seedlings were ground to a fine powder using liquid nitrogen and resuspended in 20 mL of cold EB1 buffer (0.44 mM sucrose, 10 mM Tris, pH 8.0, 10 mM $MgCl_2$, 5 mM B-Me, 0.1 mM PMSF). The solution was filtered through two layers of miracloth and centrifuged at 3,000 x g, 20 min, 4°C. The pellet was then resuspended in 1 mL of cold EB2 Buffer (Sucrose 0.25M, 10 mM Tris, pH 8.0, 10 mM $MgCl_2$, 1% v/v Triton X-100, 5 mM □-Me, 0.1 mM PMSF) and centrifuged at 4°C, 12,000 g for 10 min. The pellet was resuspended in 300 µL of cold EB3 buffer (sucrose 1.7M, 10 mM Tris, pH 8.0, 2 mM $MgCl_2$, 0.15% v/v Triton X-100, 5 mM □-Me, 0.1 mM PMSF) and the sample was overlaid on top of 300 µL of cold EB3 and centrifuged at 4°C, 16,000 x g for 1 hour. Supernatant was transferred to a low-retention tube, snap-frozen and stored at -20°C.

Chromatin immunoprecipitation: ChIP was performed as described (Deal & Henikoff, 2010) with the following modifications. Briefly, 380 ng of Drosophila chromatin (DNA) was mixed with 20 ng of Arabidopsis chromatin as a spike-in control (5%), and incubated with 1 μg of each of the following antibodies: H3 (Abcam, Cambridge MA, Catalog #1791), H3K4me3 (Abcam, Catalog #8580) and H3K36me3 (Abcam, Catalog #9050) for 12 to 18 hours at 4°C. Immunoprecipitated histone-DNA complexes were incubated with 25 μL Dynabeads protein G (ThermoFisher, Catalog #10004D) for 2 hours at 4°C, followed by 5-min washes with 1 mL Low Salt Buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% v/v SDS, 1% v/v Triton X-100, 2 mM EDTA), 1 mL High Salt Buffer (20 mM Tris, pH 8.0, 500 mM NaCl, 0.1% v/v SDS, 1% v/v Triton X-100, 2 mM EDTA), 1 mL LiCl Wash buffer (10 mM Tris, pH 8.0, 250 mM LiCl, 0.1% v/v Na-Deoxycholate, 1% v/v NP-40 substitute, 1 mM EDTA) and 1 mL TE (10 mM Tris, pH 8.0, 1 mM EDTA). Histone-DNA complexes were eluted from the magnetic beads with X-ChIP elution buffer (100 mM NaHCO$_3$, 1% v/v SDS), treated with RNAse A (ThermoFisher, Catalog #EN0531) at 37°C for 30 min and Proteinase K (ThermoFisher, Catalog #AM2546) at 55°C for 12 to 18 hours. DNA was purified with Zymo Research ChIP DNA clean & concentrator kit (Zymo Research, Catalog #D5205). Purified DNA was quantified with Qubit 1X HS DNA kit (ThermoFisher, Catalog #Q33230). Input sample fragment size was determined with TapeStation High-Sensitivity D5000 Screen tapes (Agilent, Catalog #5067-5592)

ChIP-seq library prepation: 2 ng of DNA were used for ChIP-seq libraries constructed with Tecan Ovation Ultralow V2 DNA-Seq Library Preparation Kit-Unique Dual Indexes (Tecan, Catalog #9149-A01). Following amplification, purified libraries were submitted to a round of double-size selection with AMPure XP beads (Beckman Coulter, Catalog# A63880) with a 0.61X-0.8X ratio. Libraries fragment size distribution was assessed with TapeStation High-Sensitivity D1000 Screentapes (Agilent, Catalog #5067-5584). Up to 16 libraries were pooled in one lane for paired-end 150 bp Illumina HiSeq sequencing.

1.4.7   CUT&Tag

CUT&Tag was performed using CUTANA™ CUT&Tag reagents (Epicypher, Durham NC, #15-1017, #15-1018, #13-0047) following manufacturer's "Direct-to-PCR CUT&Tag Protocol" with minor modifications: Briefly, purified nuclei were washed 1 times with cold Antibody150 buffer,

and protocol was started at Section III "Binding of Primary and Secondary antibodies" and followed as described: https://www.epicypher.com/content/documents/protocols/cutana-cut&tag-protocol.pdf, last accessed on May 15th, 2021.

1.4.8 Data processing

Raw reads were trimmed using Trimmomatic version 0.39 (64) to filter out low quality reads (Q>30) and clean adapter reads. Cleaned reads were aligned to the *Drosophila melanogaster* genome (BDGP Release 6 + ISO1 MT/dm6 from UCSC) using splicing-aware aligner STAR version 1.3 (65) for RNA-seq, and Bowtie2 version 2.3.5.1 (66) for Omni-ATAC, ChIP-seq and CUT&Tag using –sensitive settings. Samtools version 1.8 (67) was used to obtain, sort and index BAM files. For genome browser inspection as well as further analyses, bigwig files were generated by normalizing datasets to count-per-million CPM coverage tracks using deepTools version 3.1.1 (68) using --normalizeUsing CPM settings. Spearman's correlation scores were calculated using deepTools' subpackages multiBigwigSummary and plotCorrelation. Metaplots and genomic distribution heatmaps were made with deepTools' subpackages computeMatrix, plotHeatmap and plotProfile. GO term analysis was performed using R package clusterProfiler (69). Spike-in normalization: FastQ Screen version 0.13.0 (70) was used to separate reads that uniquely mapped to either the genome of *Drosophila melanogaster* (BDGP Release 6 + ISO1 MT/dm6 from UCSC) or *Arabidopsis thaliana* (Tair10 – Arabidopsis.org) using the filter option and with sensitive parameters. Each fastq file was aligned and processed separately, and alignment rates to each genome file were used to calculate spike-in factors (Orlando et al., 2014). Calculated spike-in factors were used to convert bam files into normalized bigwig files using deepTools bamCoverage subpackage, with –scaleFactor setting, generating Reference-adjusted Reads Per Million (RRPM) files with a 10-bp resolution. Encode blacklist regions were removed. Spearman correlation scores were calculated by partitioning the mappable genome into 500-bp bins and obtaining the RRPM within each bin. Omni-ATAC narrow peaks were obtained using MACS2 version 2.1.2 (71) with settings: "--nolambda --nomodel --extsize 150 --shift 75 --keep-dup all", and H3K4me3 ChIP-seq and CUT&Tag peaks were obtained with settings: "--nolambda --nomodel --keep-dup all". FRiP scores were calculated using FeatureCounts of Subread version 1.6.1 (72). Peak overlap and genomic distribution of peaks was determined using R package ChIPseeker (73).

1.5　Figures
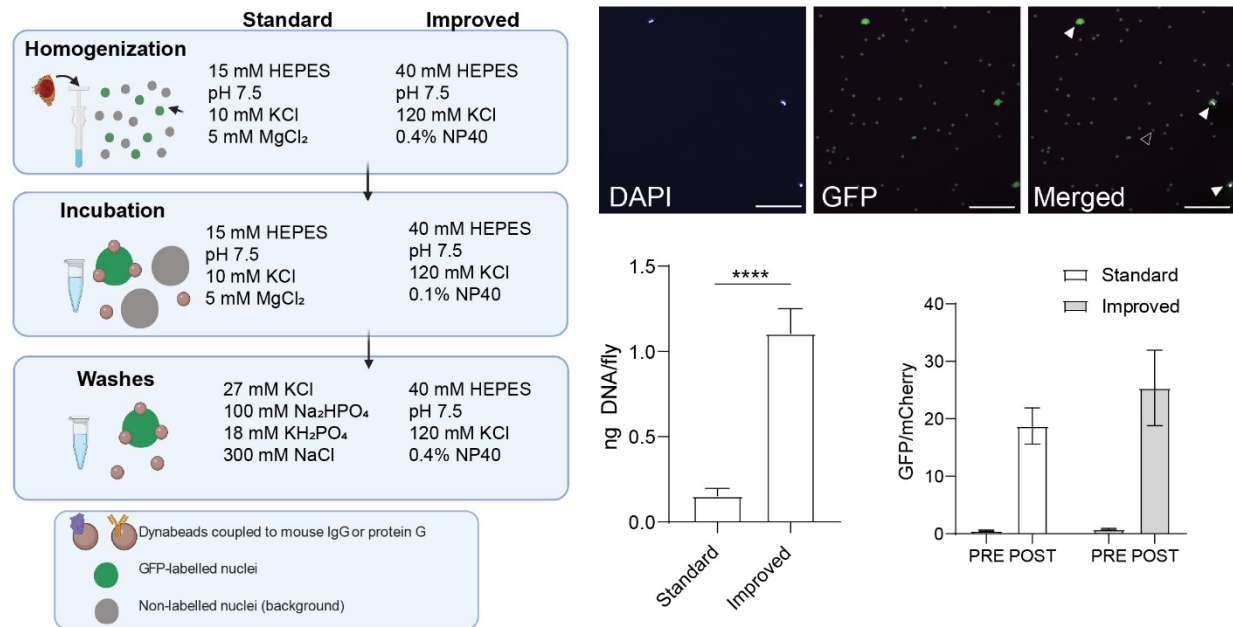
Figure 1.1 Schematic diagram depicting the nuclear immuno-enrichment (NIE) protocol highlighting major differences in buffer composition between the 'standard' and 'improved' methods. Heads from flies expressing Rh1>GFPKASH were homogenized, followed by bead-antibody incubation and washes. B. Microscopy images of POST sample using the 'improved' method. Scale bars: 50 μM. White arrowhead: bead-bound nuclei. Black arrowhead: single bead. C. Bar plot showing DNA yields when Rh1>GFPKASH nuclei were enriched using either the 'standard' or 'improved' NIE method (mean ± standard deviation (SD), n=4, p-value t-test). D. Bar plot showing qPCR enrichment for GFP and mCherry in the PRE and POST-NIE samples comparing 'methods (mean ± SD; n=3, p-value t-test).
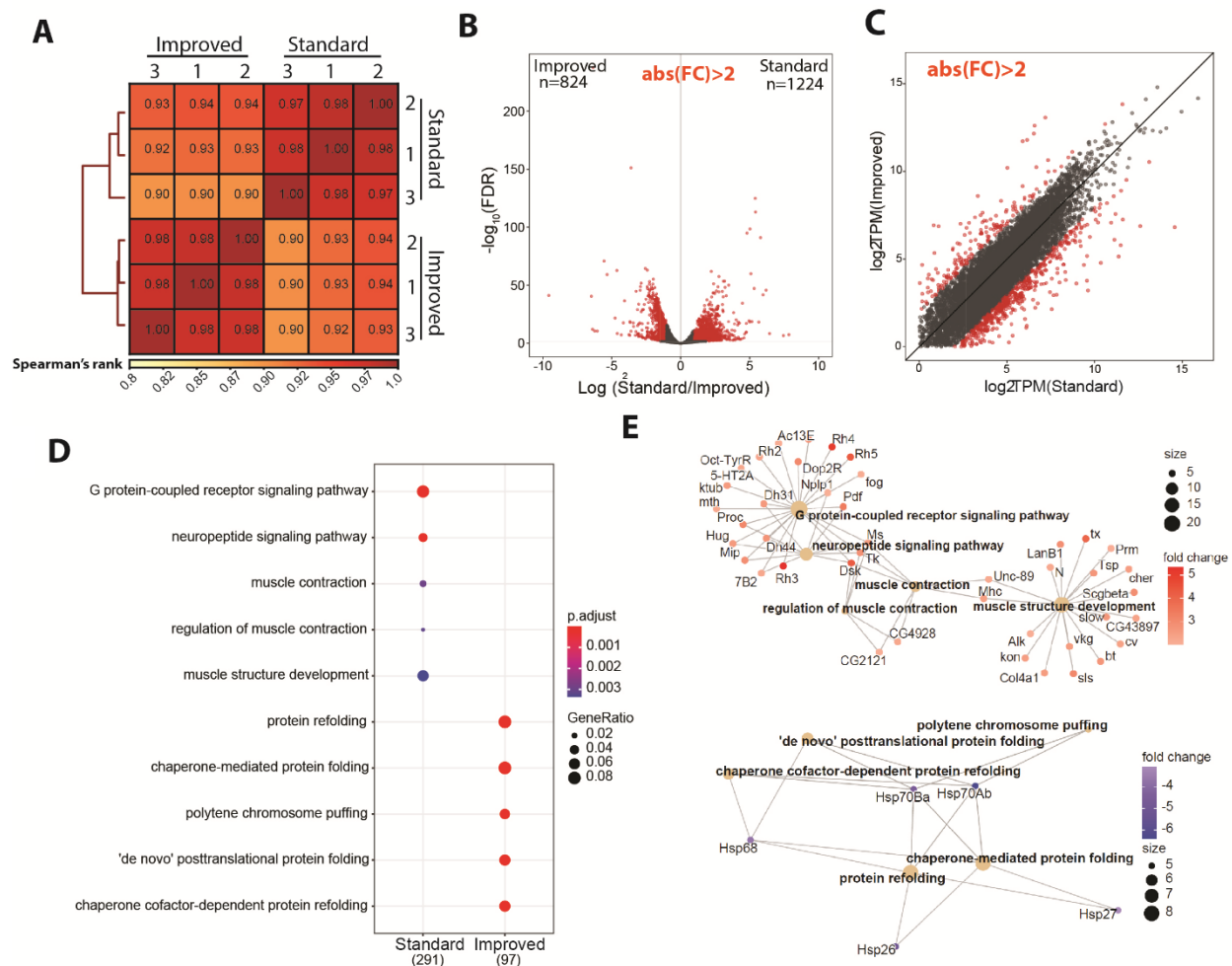
Figure 1.2 A. Spearman correlation heatmap of gene expression profiles from nuclear RNA-seq of nuclei extracted with standard and improved method (n=4). Scores between 0 and 1 shown in each box correspond to Spearman's rank score. B. Volcano plot showing the differentially expressed genes between methods. Genes with significant differential expression (FC > 2, FDR < 0.01) are highlighted in red. C. Scatter plot showing $\log_2$-transformed transcript per million (TPM) values between methods. DEGs highlighted in red, as in panel B. D. Gene Ontology (GO) term analysis on genes that are overrepresented in either the 'standard' or 'improved' method. E. Gene Concept Network plot (Cnetplot) highlighting linkage of individual genes and associated functional categories of genes over-represented in standard (top) and improved (bottom) dataset. Color intensity represents fold change between conditions.

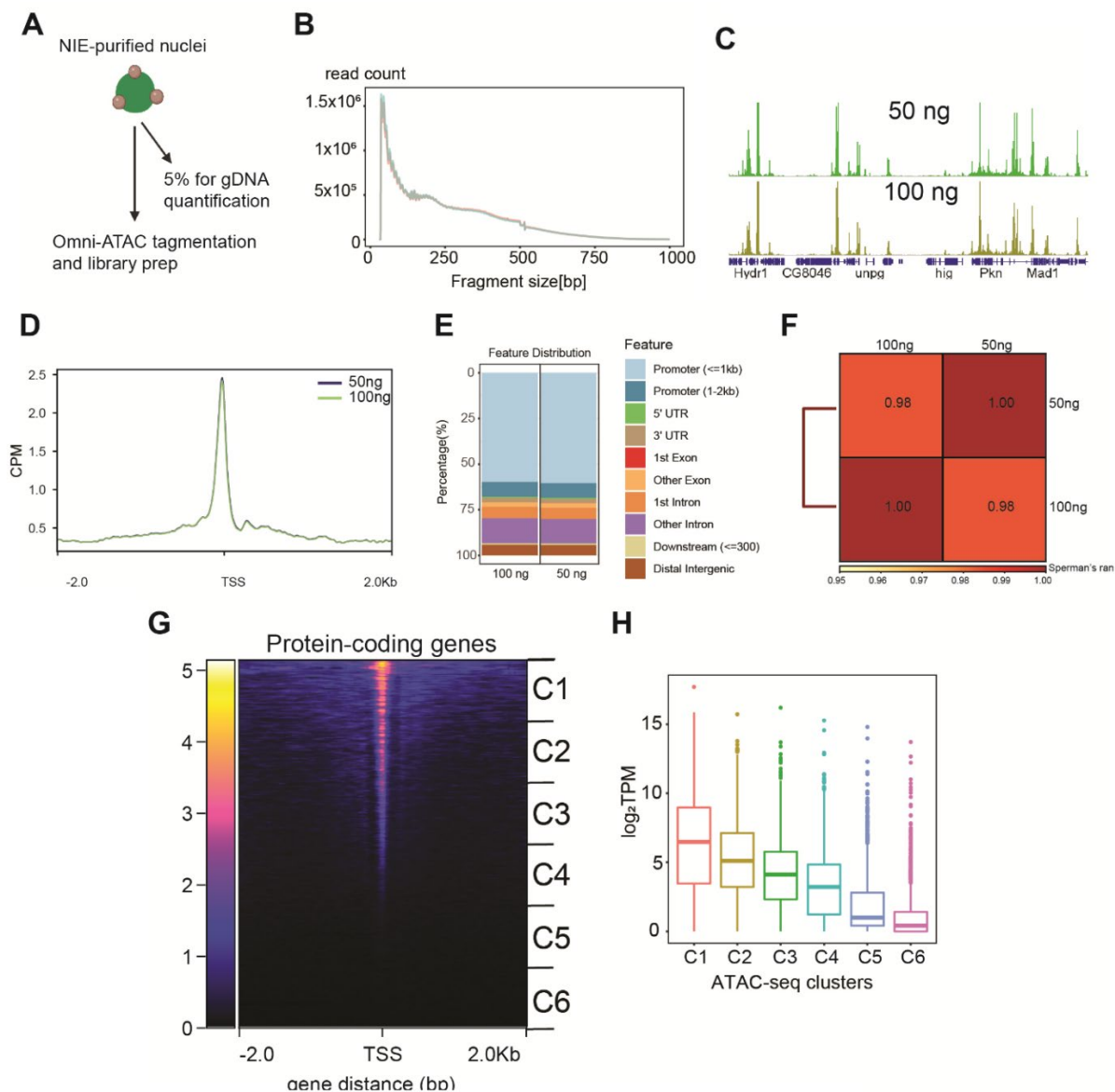Figure 1.3 A. Diagram depicting Omni-ATAC approach applied to NIE-purified nuclei. After NIE purification, a fraction of nuclei is used for genomic DNA extraction and quantification to determine the input material for Omni-ATAC. Nuclei remain on ice until tagmentation, followed by two washes with tagmentation buffer without Tn5 enzyme. Upon washes, nuclei are tagmented using standard ATAC-seq conditions. B. Fragment size distribution of Omni-ATAC libraries using 50 ng (light green) or 100 ng (light red) as starting material. C. Genome browser views of counts per million (CPM)-normalized Omni-ATAC signal with genes shown in blue. D. Metaplot of CPM-normalized Omni-ATAC signal around the transcription start site (TSS) averaged for all protein-coding genes in the 50 ng and 100 ng samples. E. Genomic distribution of accessible peaks of 50 ng- and 100 ng- associated dataset. F. Spearman correlation heatmap of Omni-ATAC read distribution over binned genome. Scores between 0 and 1 shown in each box correspond to Spearman's rank score. G. Heatmap showing CPM-normalized Omni-ATAC signal around TSS of protein-coding genes of 100 ng-associated dataset. Clusters used for transcript boxplot are highlighted. H. Boxplot showing $\log_2$-transformed TPM scores for each cluster defined in 3G.
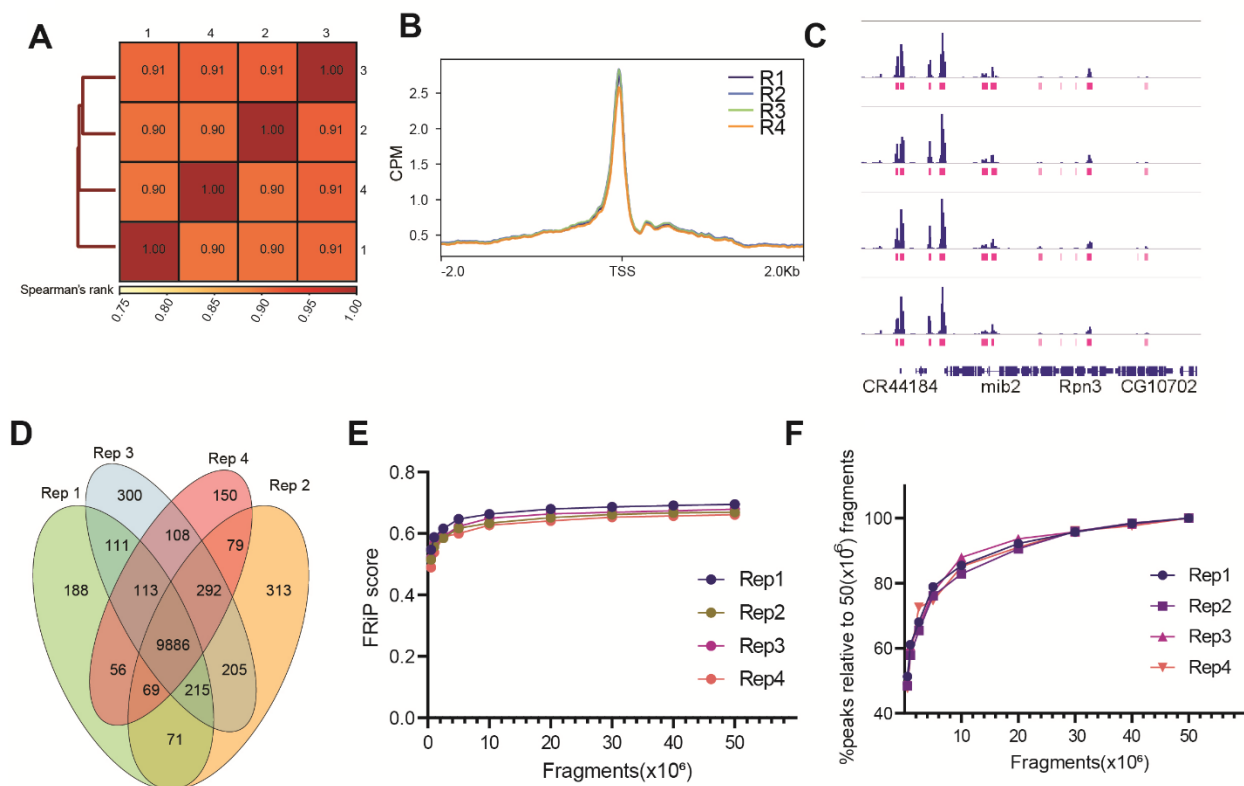
Figure 1.4 A. Spearman correlation heatmap of Omni-ATAC read distribution over binned genome. Scores between 0 and 1 shown in each box correspond to Spearman's rank score. B. Metaplot of CPM-normalized Omni-ATAC signal around TSS averaged for all protein-coding genes comparing replicates (n=4). C. Genome browser inspection of CPM-normalized Omni-ATAC signal for each replicate, coupled with narrow peaks (pink). Genes are shown in blue. D. Venn diagram showing peak overlap/similarity between replicates. E. Fraction of Reads in Peaks (FRiP) scores of Omni-ATAC peaks comparing replicates down-sampled from 0.5 to 50 million mapped fragments. F. Percentage of peaks called relative to peaks called using the Omni-ATAC replicate #1, with 50×10^6 mapped fragments as absolute percent of peaks.

Figure 1.5 A. Diagram depicting Chromatin Immunoprecipitation (ChIP)-seq approach coupled to NIE-purified nuclei. Before adding the ChIP antibody, a fraction of soluble Drosophila chromatin (input) is quantified, to adjust final amount of chromatin per replicate, as well as to define amount of spike-in genome (In this case, 5% of Arabidopsis chromatin). B. Metaplots of H3 (dark blue), H3K4me3 (light blue) and H3K36me3 (yellow) ChIP-seq signal (CPM) over gene bodies averaged for all protein-codin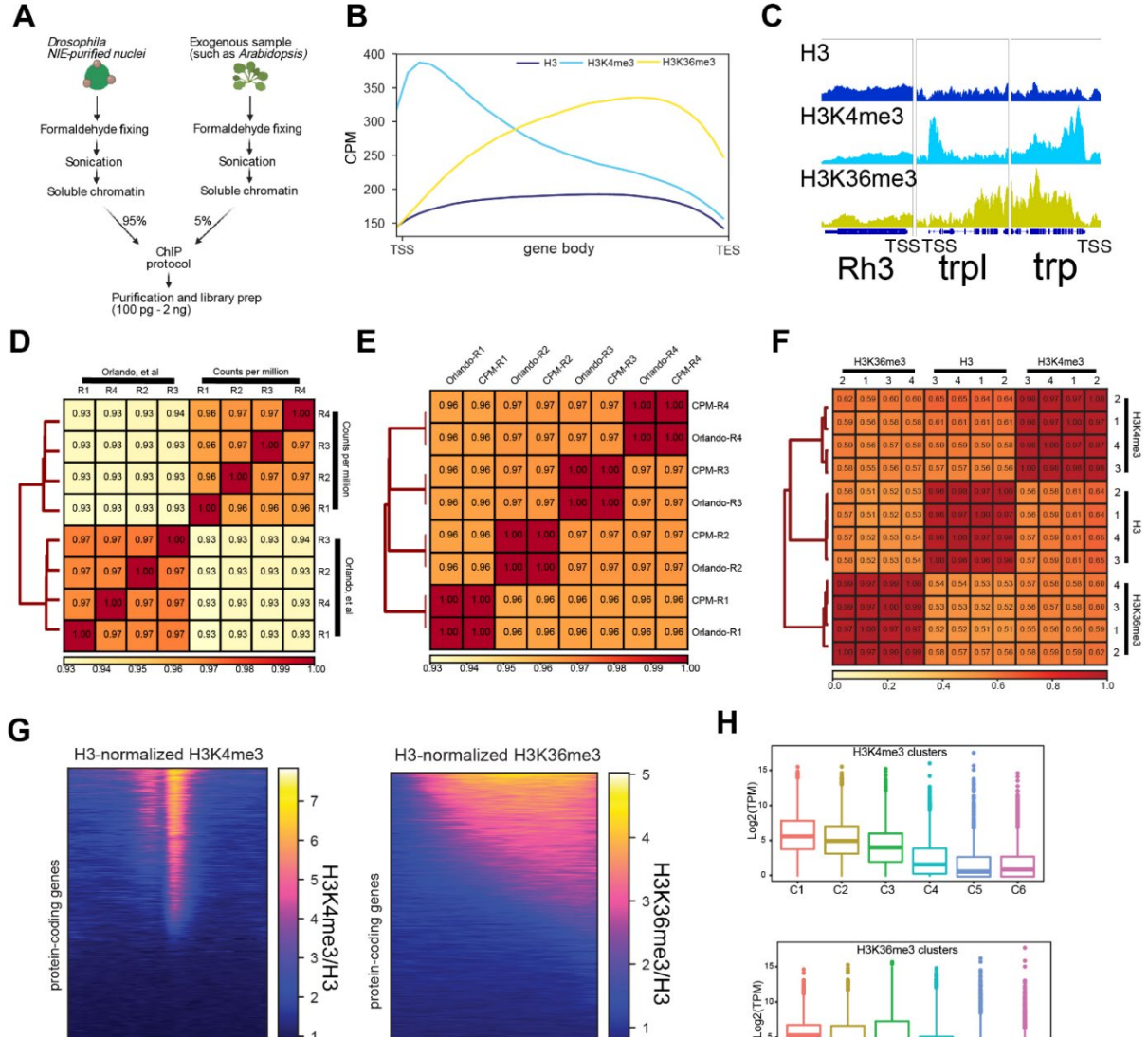g genes. C. Genome browser inspection of H3, H3K4me3 and H3K36me3 ChIP-seq signal(CPM) around the inner photoreceptor-specific gene *Rh3*, which is not expressed in outer photoreceptors, and two highly expressed outer photoreceptor-specific genes *trp* and *trpl*. D. Spearman correlation heatmap of H3K4me3 ChIP-seq data comparing Spike-in and CPM normalization. Spearman's rank scores are based on read distribution over binned genome. E. Spearman correlation heatmap of H3K36me3 ChIP-seq data comparing Spike-in and CPM normalization. Spearman's rank scores are based on read distribution over binned genome. F. Spearman correlation heatmap of reads that align to binned genome for all replicates of H3, H3K4me3 and H3K36me3 ChIP-seq datasets. G. Heatmap showing signal for all protein coding genes of H3-normalized H3K4me3 (left) and H3-normalized H3K36me3 (right). F. Boxplots showing transcript level expressions of H3K4me3 (top) or H3K36me3 clusters (bottom).
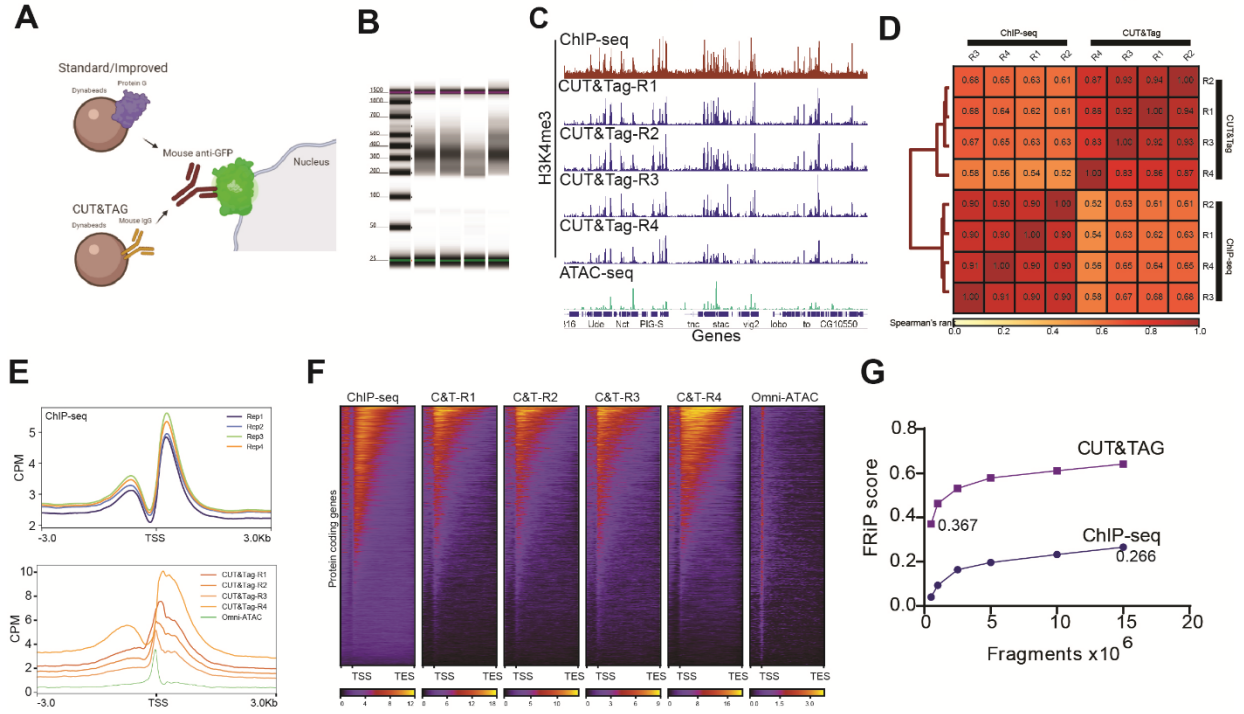
Figure 1.6 Schematic diagram representing the major difference between bead-antibody conjugation necessary to perform CUT&Tag in NIE-purified nuclei. Protein-G Dynabeads recognize both rabbit and mouse antibodies, while Mouse Pan IgG Dynabeads only recognize mouse antibodies. Nuclei preparation contains excess Dynabeads, therefore the protein G can interfere with CUT&Tag because it can bind the rabbit antibodies used to tag chromatin targets, such as H3K4me3. B. Tape Station profiles of H3K4me3 CUT&Tag libraries. C. Genome browser inspection (IGV) of CPM-normalized H3K4me3 ChIP-seq (top), H3K4me3 CUT&Tag replicates (medium) and Omni-ATAC (bottom). All samples were obtained from 10-day old male flies. Genes are shown in blue. D. FRiP score comparison between H3K4me3 CUT&Tag replicate 4 and H3K4me3 ChIP-seq replicate 1. Both samples were down sampled from 0.5 to 15 million mapped fragments. E. Metaplots of CPM-normalized H3K4me3 ChIP-seq (top) and H3K4me3 CUT&Tag (bottom) (n = 4 for each method). F. Heatmaps showing CPM-normalized H3K4me3 ChIP-seq (left-most) and H3K4me3 CUT&Tag signal for all replicates, with rows representing the same gene across all heatmaps. G. Spearman correlation heatmap of read distribution over H3K4me3 peaks called using ChIP-seq datasets. Correlation was calculated for H3K4me3 ChIP-seq and CUT&Tag replicates.

A

Dynabeads · Protein G · Mouse anti-GFP

Compatible with

nuclear RNA-seq
Omni-ATAC/ATAC-seq
ChIP-seq

Dynabeads · Mouse IgG · Mouse anti-GFP

CUT&Tag
nuclear RNA-seq
Omni-ATAC/ATAC-seq
ChIP-seq
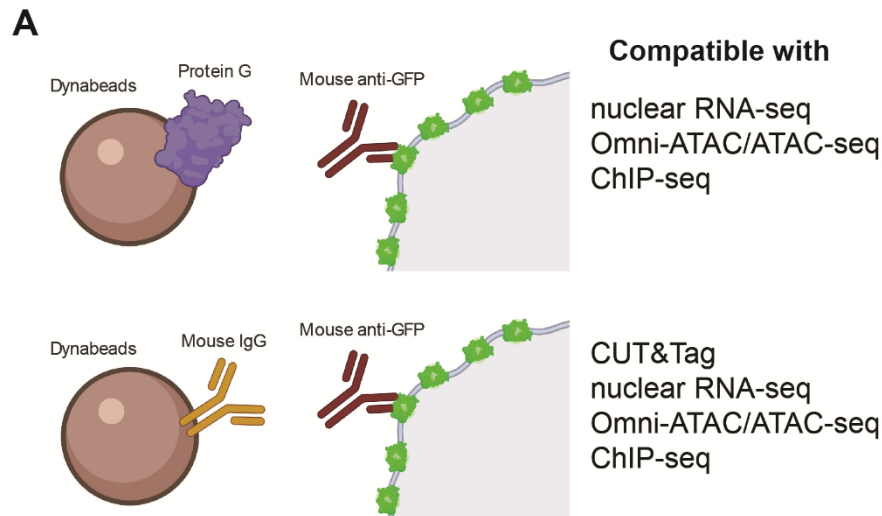
B

Technique	# nuclei	Starting material for one biological replicate
RNA-seq	~50,000	10 ng RNA
Omni-ATAC	~125,000	50 ng gDNA
ChIP-seq	~1×10^6	400 ng gDNA (3 antibodies + 5% input)
CUT&Tag	~250,000	100 ng gDNA

C

Fly line	Chr	Insertion site	Exp. System	BDSC stock number
UAS-GFP[KASH]	2	attP40	Gal4-UAS	#92582
UAS-GFP[KASH]	3	attP2	Gal4-UAS	#92580
UAS-mCherry-2xFLAG[KASH]	2	attP1:55C4	Gal4-UAS	#92585
UAS-6xmyc[KASH]	3	Su(Hw)attP2	Gal4-UAS	#92581
QUAS-GFP[KASH]	2	attP40	QF-QUAS	#92583
QUAS-GFP[KASH]	3	attP2	QF-QUAS	#92584

Figure 1.7 A. Schematic diagram representing the two versions of the "improved" NEI-method. The first version (top) uses protein G-coupled magnetic Dynabeads, and can be coupled with RNA-seq, Omni-ATAC and ChIP-seq. The second version (bottom) uses Mouse IgG-coupled magnetic beads, and can be coupled with CUT&Tag, RNA-seq, Omni-ATAC and ChIP-seq. B. Table describing the available fly lines to perform NIE either using the Gal4-UAS or the QF-QUAS system.

CHAPTER 2.   THE CLOCK: CYCLE COMPLEX IS A MAJOR TRANSCRIPTIONAL REGULATOR OF *DROSOPHILA* PHOTORECEPTORS THAT PROTECTS THE EYE FROM RETINAL DEGENERATION AND OXIDATIVE STRESS

Declaration of collaborative work

The work described in this chapter was the collaborative work of Juan Jauregui-Lozano, Hana Hall, Sarah Stanhope, Kimaya Bakhle, and Makayla Marlin, under the supervision of Vikki M. Weake. Hana Hall performed the light-dark RNA-seq. Sarah Stanhope performed the Gluthathione measurements. Kimaya Bakhle performed some of the optic neutralization, and performed the blind scoring of microscopy images. Makayla Marlin performed some quantitative PCR. All other experiments, as well as data visualization and study conceptualization were performed by Juan Jauregui-Lozano. Juan. The original draft was written by Juan Jauregui-Lozano, and edited by Juan Jauregui-Lozano and Vikki Weake

2.1   Introduction

One of the hallmarks of the aging eye, as well as many age-related eye diseases, is the loss of photoreceptor function and survival (6,7). The aging epigenome and transcriptome of cells in the retina undergo changes that correlate with decreased visual function and increased cell death (13,74,75). Importantly, disruption of epigenetic mechanisms is associated with the onset of age-related eye diseases, such as age-related macular degeneration (76,77), suggesting that transcriptional regulation contributes to the changes in homeostasis that are observed in the aging eye. However, we still have only a basic understanding of how the molecular mechanisms that drive the age-associated changes in the transcriptome increase the risk of ocular disease with advanced age.

Transcription factors (TF) function as regulatory hubs of gene expression programs in a highly-tissue specific manner. While several conserved pathways contribute to changes in the aging transcriptome across tissues, such as the longevity-associated FOXO and Insulin axis (78–80), age-associated changes in gene regulatory networks can be highly specific to individual cell types (3,75). The emergence of bioinformatic and computational approaches that combine chromatin

accessibility data with transcription factor binding sites (TFBS) has allowed researchers to interrogate unbiasedly how transcription factor activity changes at a genome-wide scale in different biological conditions by estimating changes in chromatin accessibility around TFBS (see (81) for review on chromatin accessibility data analysis). The majority of chromatin accessibility studies in the eye have focused on development in invertebrate (60) or vertebrate models (82), and TF activity in the developing eye has often been assessed based on transcriptomic-based co-expression inference rather than chromatin accessibility (83). Further, ATAC-seq analysis of retinal pigmented epithelium from patients with age-related macular degeneration identified global changes in chromatin accessibility at the onset of the disease state, suggesting that differential activity of regulatory elements strongly contributes to the initial stages of age-associated ocular disease (84). Thus, identification of TFs with differential activity in the aging eye could provide insight into the mechanisms that contribute to the increased risk of retinal degeneration during aging.

Drosophila melanogaster, like humans, experience an age-dependent decline in visual function coupled with increased risk of retinal degeneration (10,13). Although flies possess a compound eye that differs anatomically from the vertebrate eye, much of the phototransduction machinery is conserved between flies and mammals, with the outer fly photoreceptors resembling vertebrate rods in function (85,86). There are six outer photoreceptors (R1 – R6 cells) within each ommatidium that express the light-sensing protein Rhodopsin 1 (Rh1) and are responsible for motion vision and phototaxis, which decline with age (3,18,19). Considering its relatively short lifespan, *Drosophila* provides a useful model system for studying the processes involved in normal aging within specific tissues, such as the eye (26). To profile the transcriptome and epigenome of specific cell types in the eye, we have developed a cell type-specific nuclei immunoenrichment technique that we have previously used to examine gene expression in aging photoreceptors (12,13,87). Here, we applied this technique to profile the transcriptome and chromatin accessibility landscapes of *Drosophila* photoreceptors across an extended time course into relatively old age. By integrating these aging data from photoreceptors, we identified 57 TF motifs that were differentially accessible during aging, suggesting age-dependent changes in TF activity. Amongst these "age-regulated TFs", we identified the binding motif of the master circadian regulators, Clock (Clk) and Cycle (Cyc), which showed sustained increases in activity with age. Using a dominant negative mutant of Clock (Clk[DN]) that disrupts the Clk:Cyc complex and abolishes

rhythmic transcription, we showed that the Clk:Cyc complex has an integral role in controlling gene expression of 15-20% of active genes, and maintaining global levels of chromatin accessibility in photoreceptors. Further, we show that expression of Clk[DN] in photoreceptors leads to progressive retinal degeneration, which was suppressed when flies were reared in constant darkness. Our data identify a novel neuroprotective role for the circadian clock in the *Drosophila* eye, and suggest that this role may become increasingly critical in advanced age to prevent retinal degeneration.

2.2 Results

2.2.1 Tissue-specific profiling of the photoreceptor nuclear transcriptome reveals significant changes during early and late aging

To profile the transcriptome of aging *Drosophila* photoreceptors, we used our recently improved tissue-specific nuclei immunoenrichment approach (87). Briefly, we tag nuclei using a Green Fluorescent Protein (GFP) fused to the "Klarsicht, ANC-1, Syne Homology" (KASH) domain, which anchors GFP to the outer nuclear membrane and allows for nuclei purification; GFP[KASH] is expressed in outer photoreceptors under Rh1-Gal4 control, herein referred as Rh1>GFP[KASH] (Fig.1A). Similar to wild-type strains, such as OregonR and Canton-S (88), Rh1>GFP[KASH] flies begin to show a substantial decline in viability after day 60, with 50% of male flies dying by day 70 (13). Thus, to generate a comprehensive aging dataset, we purified photoreceptor nuclei over 10-day windows from day 10 until day 60 and performed RNA-seq using 400 age-matched male flies per biological replicate (Fig. 1B).

We assessed the overall variability of the RNA-seq samples using Principal component analysis (PCA). The RNA-seq samples clustered together with 48.3% of the variation separating the samples by age in a progressive manner (Fig 1C). We also observed a similar grouping by age using Pearson's correlation analysis with high concordance between the three biological replicates (Pearson's $r > 0.91$) (S1 Fig). These observations reveal progressive changes in transcription during aging in photoreceptors, similar to previous studies from our group that aged flies to day 40 (D40) (13). Differential expression analysis of each time point relative to day 10 (D10), the youngest state, revealed significant changes in gene expression during aging (p-adj<0.05) (Fig. 1D-S1 table). Hierarchical clustering of the genes that were differentially expressed between any

age and D10 revealed that 1412 and 982 genes showed progressive increases, or decreases in expression, respectively. Gene Ontology (GO) term analysis revealed that the genes that were age-upregulated (Clusters 2 and 3) were enriched for several processes including cytoplasmic translation and double strand break repair (Fig. 1E). In contrast, the genes that were age-downregulated (Cluster 5) were enriched for neuronal processes, such as cognition and regulation of nervous system development. Although 75% of the differentially expressed genes (2394 out of 3156) exhibit progressive changes, we also identified two clusters with expression changes only during early aging (D20; Clusters 1 and 4). We note that while D20 was separated by age from the other samples along PC1, it was also separated from its nearest time points (D10 and D30) along PC2 (Fig. 1C), consistent with these early and transient changes in gene expression in aging photoreceptors. Interestingly, Cluster 4, which contained genes that showed an early decrease in transcription but increased levels again at later time points, was enriched for protein folding. These age-associated transcriptomic signatures are consistent with our previous observations in aging photoreceptors and with aging studies from other tissues (13,74).

2.2.2 Photoreceptors undergo substantial changes in transcription factor activity during aging

To identify mechanisms that contribute to the age-associated transcriptomic changes in photoreceptors, we sought to evaluate changes in TF activity. TFs regulate transcriptional outputs of their targets by acting as activators or repressors of gene expression at the transcriptional level. Importantly, TF activity can be affected by several factors, including protein levels, localization, and post-translational modifications (89,90). Because of the technical complexity of isolating intact photoreceptors for proteomic studies, and the relatively low protein abundance of TFs in the eye (91), we used our RNA-seq data to assess how transcript levels of genes associated with the GO category "DNA-binding transcription factor activity, RNA polymerase II-specific" (GO:0000981) changed during aging. Notably, 23% of TFs showed significant differential expression during aging at the nuclear transcript level (Fig. 2A), suggesting that differential TF activity could contribute to the aging transcriptome of photoreceptors. To identify TFs with differential activity during aging, we used *diffTF*, a software package that estimates genome-wide changes in TF motif/binding activity based on differences in aggregate ATAC-seq signal around predicted/validated TF binding sites, or TFBS (92). We refer to TF motif/binding activity as "TF

activity", as defined in the *diffTF* study (27). Although *diffTF* analysis provides an estimate of TF activity, rather than a direct measurement of TF binding to DNA, *diffTF* was highly ranked as an approach to assay genome-wide changes in TF activity (93). Thus, we first profiled the chromatin accessibility landscape of aging photoreceptors at the same time-points as for RNA-seq. PCA of ATAC-seq samples revealed that 69.7% of the variation could be explained by age (Fig. 2B), and Pearson correlation analysis showed high concordance between biological replicates. Additionally, accessible peak annotation revealed a stable distribution of peaks through-out aging (S2A Fig). Together with the RNA-seq aging time course, these data indicate that the chromatin accessibility and transcriptional landscape of photoreceptors undergo progressive changes during aging. We then used *diffTF* to compare all time points relative to D10. The TFBS used for this analysis were generated by CIS-BP, which provides a comprehensive dataset of experimentally validated TFBS (94). We also examined *de novo* and known motifs identified using Homer (95) (Fig. 2C). We took advantage of the aging RNA-seq time series data to perform *diffTF* analysis in integrative mode, enabling us to discard TFs that were not expressed in adult photoreceptors. Using this approach, we identified 57 TFs whose binding sites showed significant differential activity during aging (FDR<0.001) (Fig. 2C), herein referred to as "age-regulated TFBSs". Most of these age-regulated TFs showed continuous changes in activity with age, with nearly two thirds showing increased activity with age. We observed an almost two-fold increase in the number of age-regulated TFs identified at D50 and D60 relative to younger ages, suggesting that late aging is associated with distinct changes in gene regulatory networks (Fig. 2D). Hierarchical clustering of these age-regulated TFs by the mean weighted difference between each age comparison resulted in grouping of TFs that are known to associate *in vivo*, such as Motif 1 Binding Protein (M1BP), DNA replication element factor (Dref) and Boundary element-associated factor of 32kD (BEAF32), which bind topologically associating domains (Ramirez, 2018). We refer to TF proteins here using their gene name and symbol (non-italicized), and provide a complete list of all genes/proteins referred to in this study with their corresponding Flybase ID numbers in Table S3. We also identified Moira (mor), which physically interacts with Similar (sima) (96). Additionally, we identified Seven up (svp) and PvuII-PstI homology 13 (Pph13), which physically interact to regulate eye-specific transcriptional programs together with Eyeless (ey) (97). Considering that svp, ey, and Pph13 show modest but significant increases in TF activity with age, our data also indicate that photoreceptor identity remains distinct even at advanced age in flies.

The stress response is transcriptionally regulated during aging across a broad range of tissues (98). Accordingly, we identified several TFs that are involved in the regulation of stress response genes as having differential activity with age including the *Drosophila* JUN-FOS complex that is formed by Jun-related antigen (Jra) and Kayak (kay) (99), the *Drosophila* homolog of Hypoxia response factor HIF-1a, sima (100), Heat shock factor (Hsf) (101), and Activating transcription factor-2 (Atf-2) (102,103). We also identified Cropped (crp), which has previously been associated with aging through *in silico* analysis due to interactions with DNA repair pathways (104).

One of the most interesting changes in TF activity during aging was in the activity of the master circadian regulators Clock (Clk) and Cycle (Cyc), which showed a progressive increase in TF activity during aging. Clk and Cyc bind the same CACGTG motif upon heterodimerization, explaining the identification and clustering of both factors in the *diffTF* analysis, which uses DNA sequence motifs as the TFBS source (S2B Fig). The circadian clock is a molecular time keeper that controls rhythmic behaviors, such as locomotor behavior, which is coordinated by pacemaker neurons located in the brain (105). In *Drosophila,* many peripheral tissues also contain working clocks, including the fat body, Malpighian tubule cells, and the retina (106). Importantly, aging has been associated with changes in the circadian clock (107,108). We note that flies for the aging time course experiment were raised under a 12:12 hour light-dark cycle, and were harvested at relative Zeitgeber time (ZT) 6 +/- 1 hour (see Methods), suggesting that the enrichment of Clk and Cyc in the *diffTF* analysis is not simply due to a time-of-day bias in sample collection. Taken together, our data identify a subset of TFs that exhibit significant changes in activity during aging in photoreceptors, including TFs associated with stress response and circadian rhythm.

2.2.3 Clock regulates the transcriptional output of phototransduction genes in photoreceptors

Clk and Cyc are basic Helix-Loop-Helix (bHLH)-TFs that form a heterodimer (Clk:Cyc) to control rhythmic expression of their targets. Canonical transcriptional targets of the Clk:Cyc complex include core clock genes, such as *vrille* (*vri*), *PAR-domain protein 1* (*Pdp1*), *timeless* (*tim*), *period* (*per*)*,* and *clockwork orange* (*cwo*). However, Clk and Cyc also regulate transcription of many other genes either directly or indirectly, and these genes are often referred as "Clock-output genes", and can include tissue-specific genes (109,110). To evaluate the biological role of Clock-

dependent transcription specifically in photoreceptors, we disrupted the Clk:Cyc complex by expressing a dominant negative version of Clk (Clk^{DN}) specifically in outer photoreceptors under Rh1-Gal4 control. Clk^{DN} lacks the basic DNA binding domain, impairing its recruitment to DNA without disrupting its ability to form a heterodimer with Cyc (Fig. 3A) (111). Expression of Clk^{DN} inhibits Clock-dependent transcription and rhythmic behaviors *in vivo* when expressed in pacemaker or antennal neurons (111). To facilitate nuclei immuno-enrichment for RNA-seq analysis, we co-expressed UAS-GFP^{KASH}. As a control, we profiled the transcriptome of photoreceptors that expressed LacZ, herein referred as Rh1>Ctrl. We collected both Rh1>Clk^{DN} and Rh1>Ctrl flies at ZT 9, when Clock-dependent transcription is active (112,113), harvesting flies at D1 and D10 to study the progressive effect of disrupting the circadian clock in adult photoreceptors (Fig. 3B). We note that Rh1-Gal4 activity begins in the very late stages of pupal development (45); thus Rh1>Clk^{DN} flies have a disrupted Clk:Cyc complex in adult photoreceptors, but not in the developing eye.

Differential gene expression (DGE) analysis of Rh1>Clk^{DN} relative to Rh1>Ctrl at either D1 or D10 revealed a consistent downregulation of direct Clk:Cyc targets such as *vri, per, tim* and *Pdp1* (Fig. 3C-S2A Fig). We also observed a significant upregulation of *Clk* itself, which we showed by qPCR reflects expression of *Clk^{DN}* rather than the endogenous wild-type *Clk* (S2B Fig). When we compared the control D1 and D10 flies, we did not observe differential expression of core clock genes (i.e. *vri, tim, per*), suggesting there is little change in the circadian clock in the early stages of adult life. In addition, only 147 genes changed in control flies between D1 and D10, indicating there is relatively little change in gene expression in general at these early stages of adult life in photoreceptors. In contrast, expression of Clk^{DN} led to significant changes in expression of 15% and 22% of actively transcribed genes in photoreceptor at D1 and D10, respectively. These data demonstrate that continued expression of Clk^{DN} leads to progressive dysregulation of gene expression in photoreceptors at nearly a quarter of expressed genes.

To further identify direct versus indirect targets of the Clk:Cyc complex, we compared previously published high-confidence binding sites identified for Clk and Cyc using ChIP-seq in *Drosophila* heads and bodies (114) with our list of accessible peaks obtained with ATAC-seq data from D10 photoreceptors. Genomic overlap analysis revealed that 10% of accessible peaks contained an experimentally identified Clk:Cyc TFBS (Fig. 3D). Next, we annotated these photoreceptor-

specific TFBS (n=1461) to their nearest transcription start site (TSS) and found that 1004 photoreceptor-expressed genes contained a potential Clk:Cyc binding site. We refer to these 1004 genes as "Clock-bound genes" or CBGs (Fig. 3E), which we compared to genes that were differentially expressed in Rh1>Clk^{DN} relative to Rh1>Ctrl at either D1 or D10. We reasoned that if a gene is bound by Clk:Cyc and is differentially expressed, then it can be classified as a "direct target" of Clk:Cyc. Using these criteria, we identified 475 direct Clk:Cyc targets in photoreceptors. In contrast, we identified a far greater number of genes (2253) that are likely to be indirect targets of Clk:Cyc regulation, at least based on the available ChIP-seq data from whole heads (114). Consistent with the predicted role of Clk:Cyc as transcriptional activators, Clk:Cyc direct targets were more likely to be downregulated relative to indirect targets upon expression of Clk^{DN} (S3C Fig). GO term analysis of direct Clk:Cyc targets showed significant enrichment of genes associated with several biological processes, including phototransduction and circadian rhythm (Fig. 3F). However, both direct and indirect Clk:Cyc targets were enriched for processes including TF activity and eye development. This is consistent with previous reports that have proposed that the Clk:Cyc complex acts at the top of a TF hierarchy, directly regulating transcript levels of key TFs in specific cell types, thereby indirectly regulating expression of their target genes (114,115). Notably, individual inspection of the TFs that were classified as direct Clk:Cyc targets confirmed that several eye-specific TFs such as *ocelli* (*oc/otd*) and *eyeless* (ey) are bound at their promoters directly by Clk and Cyc (Fig. 3G). To further identify the magnitude of changes in expression of these pathways upon disruption of Clk:Cyc, we performed gene set enrichment analysis comparing Clk^{DN} vs Control at either D1 or D10 and obtained ridge plots showing the density of fold change for the genes associated with each pathway. Whereas upregulated genes showed significant enrichment of several biological processes, including TF activity, downregulated genes were associated with several phototransduction-associated pathways, including deactivation of rhodopsin signaling (Fig. 3H, S3D-E Fig).

Because light entrains the circadian clock to activate Clock-mediated transcription, we next performed RNA-seq in photoreceptors from Rh1>GFP^{KASH} flies reared in constant darkness (DD). Under these conditions, we observed differential expression of genes associated with several biological processes, including phototransduction and circadian rhythm in dark raised flies relative to LD (S3F-G Fig). Because these processes were also enriched in the genes with differential

expression upon disruption of Clk:Cyc activity, our data suggest that these biological processes are normally regulated by the circadian clock in photoreceptors in response to the light/dark cycle.

Taken together, these data show that the Clk:Cyc complex is a major transcriptional regulator of the photoreceptor transcriptome, including the phototransduction pathway. Additionally, our data suggest that Clk activity regulates gene regulatory networks by regulating expression of TFs that in turn direct expression of a large proportion of the transcriptome in photoreceptors.

2.2.4 Expression of Clk^{DN} leads to genome-wide changes in TF activity and a widespread decrease in chromatin accessibility

Because expression of Clk^{DN} led to transcriptional dysregulation of TFs, we further investigated the TFs that showed changes in transcript level upon disruption of the Clk:Cyc complex. Gene Concept Network (cnet) plots demonstrated significant enrichment of 62 TFs amongst the upregulated genes in flies expressing Clk^{DN} (Fig. 4A). Based on these changes in TF expression, we wondered if expression of Clk^{DN} would also lead to significant changes in TF activity relative to Rh1>Ctrl flies. To test this, we profiled chromatin accessibility in Rh1>Clk^{DN} versus control, and integrated the RNA-seq and ATAC-seq data using the *diffTF* approach, as in Fig. 2. PCA analysis of the accessible chromatin landscape revealed that 77% of the variation was explained by expression of Clk^{DN} (Fig. 4B). In addition, Clk:Cyc targets, such as *tim, per, vri* and *Pdp1,* exhibited significant decreases in accessibility through-out their gene bodies upon expression of Clk^{DN} (Fig. 4C). Importantly, genomic annotation of accessible peaks revealed a stable distribution of peaks in all samples (S4A Fig), suggesting that disruption of the Clk:Cyc complex does not lead to overall changes in the genome-wide distribution of accessible peaks. Rather, these data show that the Clk:Cyc complex promotes chromatin accessibility at target genes, consistent with the well characterized role of Clk:Cyc in transcription activation (116).

Using *diffTF*, we identified 37 TFs with differential activity upon expression of Clk^{DN}, or "Clock-regulated TFs". Whereas 15 TFs had decreased activity, 22 showed increased activity in Rh1>Clk^{DN} relative to Rh1>Ctrl (Fig. 4D). Interestingly, several of the genes encoding these Clock-regulated TFs are also directly bound by Clk and Cyc at their promoters (see Figure 3), including the eye-specific TFs *oc/Otd* and *ey*, as well as *Xrp1, onecut, crp,* and *Trl*. We also identified increased TF activity of the circadian regulator *vri*, which represses transcription of *Clk*

(117). Consistent with the reported role of *vri* as a repressor, we observed decreased levels of wild-type *Clk* transcript in Rh1>Clk^{DN} flies using qPCR with primers that differentiate between the wild-type and dominant negative version (see S2B Fig). Thus, these data suggest that the Clk:Cyc complex contributes to the maintenance of gene regulatory networks in photoreceptors by regulating transcript levels and/or TF motif/binding activity of many transcription factors.

Visual inspection of chromatin accessibility tracks showed that the Rh1>Clk^{DN} flies had an overall decrease in chromatin accessibility around promoters and throughout gene bodies (Fig. 4C). To further evaluate this, we obtained gene metaplots averaging the ATAC-seq signal around promoters for all actively expressed genes in Rh1>Clk^{DN} and Rh1>Ctrl flies. Strikingly, we observed that expression of Clk^{DN} led to a widespread decrease in chromatin accessibility around transcription start sites, and this decrease in global accessibility was exacerbated at D10 relative to D1 (Fig. 4E). These gene-based observations were corroborated by peak-based quantification (S4B Fig). Thus, our data suggest that disruption of the Clk:Cyc complex results in a global decrease in chromatin accessibility in photoreceptors, even though Clk and Cyc only have 475 direct gene targets in this cell type (Fig. 3E). To further gain insight if these changes in accessibility were progressive between D1 and D10 for a given gene, we obtained heatmaps of ATAC-seq fold change signal for actively expressed genes at either D1 or D10 in which both heatmaps were sorted identically (Fig. 4F). These heatmaps revealed that genes that had decreased accessibility at D1 also showed sustained decreases in accessibility at D10, suggesting that disruption of the Clk:Cyc complex has a role in promoting chromatin accessibility in photoreceptors at a large fraction of actively expressed genes, although our data does not identify if this role is direct or indirect.

Taken together, our data shows that expression of Clk^{DN} leads to dysregulation of TF levels and/or activity, suggesting the Clk:Cyc modulates gene regulatory networks associated with these TFs. In addition, expression of Clk^{DN} leads to a widespread decrease in chromatin accessibility that is independent from changes in gene expression, suggesting that the circadian clock contributes to the global maintenance of chromatin remodeling of actively expressed genes in *Drosophila* photoreceptors.

2.2.5 Disruption of Clock activity leads to light-dependent retinal degeneration and light-independent accumulation of oxidative stress

Disruption of the phototransduction machinery by mutations in phototransduction genes that leads to loss of function or decreased expression is often associated with retinal degeneration (118). Consistent with the GSEA analysis (see Fig. 3H), cnet plots evaluating RNA-seq fold change for genes associated with phototransduction revealed significant decreases in transcript levels of 22 genes in flies expressing Clk^{DN} (Fig. 5A-S5 Fig). These genes included many phototransduction components whose loss leads to light-dependent retinal degeneration, such as *Cds*, *Arr2*, *rdgB*, *trp*, *rdgC*, and *ninaC* (marked with a red asterisk, Fig. 5A) (118). Thus, we hypothesized that expression of Clk^{DN} in *Drosophila* photoreceptors would lead to retinal degeneration. To test this, we examined photoreceptor integrity using optical neutralization (see Methods) in Rh1>Clk^{DN} and Rh1>Ctrl flies that were maintained in either standard 12:12 hour light:dark (LD) or free-running, constant darkness (DD) conditions. We note that all of the flies used in this study had pigmented eyes (S5B fig), and were not susceptible to light-mediated damage as is the case for w^{1118} flies. We found that Rh1>Clk^{DN} flies presented progressive retinal degeneration starting at day 5, when raised under standard light:dark (LD) conditions relative to Rh1>Ctrl flies (Fig. 5B top and 5C-left). Importantly, photoreceptors were intact at earlier adult stages just after eclosion (D1), although qPCR analysis from 1-day old heads shows that the *Clk^{DN}* transcript is already expressed at this age (S3B Fig). We observed retinal degeneration in two independent *UAS-Clk^{DN}* lines inserted on different chromosomes, suggesting that retinal degeneration is unlikely to result from insertion position of the transgene; moreover, all experiments were performed in progeny hemizygous for the *UAS-Clk^{DN}* transgene. To test if the retinal degeneration resulting from Clk^{DN} expression was dependent on light exposure, we raised Rh1>Ctrl and Rh1>Clk^{DN} flies in constant darkness (DD) and monitored retinal degeneration. Strikingly, rearing Rh1>Clk^{DN} flies in constant darkness prevented the onset of retinal degeneration in both Clk^{DN} lines (Fig. 5B-bottom and 5C-right) demonstrating that expression of Clk^{DN} results in light-dependent retinal degeneration in adult *Drosophila* photoreceptors. Since Clk^{DN} expression resulted in light-dependent retinal degeneration, we next asked if expression of the analogous dominant negative for its partner Cycle (Cyc^{DN}) caused a similar phenotype. Surprisingly, we did not observe any retinal degeneration in Rh1>Cyc^{DN} flies reared in standard LD conditions at either D5 or D10 (S5C Fig). When we examined the relative level of *Clk* and *cyc* transcripts in photoreceptors, we found that *Clk*

transcripts are 10-times more abundant than *cyc* (S5D Fig). In addition, proteomic analysis of fly heads during the day revealed that there are about 5-fold fewer peptides corresponding to Cyc versus Clk (112). Thus, the differences in phenotype between Clk[DN] and Cyc[DN] flies could arise because of differences in protein abundance; however, it remains possible that Clk and/or Cyc have independent functions in photoreceptors outside of the canonical Clk:Cyc complex.

The circadian clock has also been shown to be necessary to protect cells against oxidative stress. For example, flies carrying the *Clk[Jrk]* allele, which produces a mis-spliced *Clk* transcript and leads to decreased clock activity (119), show increased levels of reactive oxygen species (ROS) in aging brains (120). Since light is a major source of oxidative stress in the eye (121) and Rh1>Clk[DN] showed light-dependent retinal degeneration, we hypothesized that Rh1>Clk[DN] flies exposed to light had increased oxidative stress levels relative to Rh1>Ctrl flies or flies reared in DD conditions. To test this, we performed a targeted metabolite assay to measure the ratio of reduced and oxidized glutathione (GSH:GSSG) in Rh1>Clk[DN] and Rh1>Ctrl flies reared in LD or DD conditions at D1 and D10. In this assay, a lower GSH:GSSG ratio is indicative of increased oxidative stress levels (122). We note that given the technical limitations for isolation of intact photoreceptors, we performed these targeted metabolite assays using dissected eyes. Using this approach, we found that oxidative stress levels did not show significant changes at D1 in any condition (Fig. 5D-left). Unexpectedly, we observed a significant decrease in the GSH:GSSG ratio in Rh1>Clk[DN] flies relative to Rh1>Ctrl flies raised in either LD and DD conditions at D10 (Fig. 5D-right). Thus, our data shows that Clock activity protects the *Drosophila* eye against oxidative stress. However, since expression of Clk[DN] only caused retinal degeneration in flies reared in LD, it is unlikely that the increased ROS levels are responsible for the retinal degeneration observed in Clk[DN] flies at D10.

Together, these data suggest that the disruption in expression of the phototransduction machinery in photoreceptors that lack Clock activity is likely responsible for the light-dependent retinal degeneration that we observed upon expression of Clk[DN]. When we performed GO term analysis of genes that were differentially expressed in Rh1>Clk[DN] at both D10 and D1, we observed significant upregulation of genes associated with response to unfolded protein and response to topologically incorrect protein, including many heat shock proteins and chaperones (Fig. 5E-5F). When the light-sensing Rhodopsin 1 (Rh1) is not properly inactivated, photoreceptor neurons

49

experience substantial protein misfolding and ER stress, which leads to retinal degeneration in a light-dependent manner (118). Since many of the genes that regulate Rhodopsin folding and inactivation were downregulated in Rh1>Clk^{DN} relative to Rh1>Ctrl flies, our data suggest that the circadian clock directly contributes to Rhodopsin maintenance in *Drosophila* photoreceptors. Importantly, the onset of light-dependent retinal degeneration associated with failures in Rh1 inactivation can be caused by several factors, including excessive endocytosis of Rh1 and increased Ca^{2+} influx (118,123–125). We propose that Clock activity is protective in the retina because it promotes expression of genes that contribute to proper recycling of Rh1 upon light-exposure. However, our data also uncover an additional neuroprotective role of Clock by contributing to the response to oxidative stress, which might be important for additional tissues that do not contain the phototransduction machinery. This secondary role of Clock in the cellular response to oxidative stress might become increasingly important in advanced age, since accumulation of ROS is one of the hallmarks of aging.

2.3 Discussion

The circadian clock maintains daily biological rhythms by controlling the expression of target output genes, and is highly conserved between *Drosophila* and humans (105). Since many genes are rhythmically expressed in the retina including most of the phototransduction machinery (114,126,127), these observations suggest that the circadian clock plays a role in homeostatic regulation of the *Drosophila retina*. Here, we report that retina-specific expression of a dominant negative mutant of Clk in *Drosophila* leads to progressive light-dependent retinal degeneration and oxidative stress, showing that the circadian clock is required to maintain *Drosophila* photoreceptor integrity. Importantly, this role for the circadian clock in maintaining visual health is conserved in mammals because retina-specific disruption of *BMAL1*, the mice homolog of *cyc*, accelerates the loss of cone viability and function in aging *chx10^{Cre};Bmal1$^{Fl/Fl}$* mice, which otherwise show a normal lifespan (128). Further, disruption of *BMAL1* leads to loss of spectral identity and integrity of cone cells in *Crx-Cre;Bmal1$^{Fl/Fl}$* (129), and the rat retina shows a circadian-dependent loss of resistance to light-mediated damage (130,131). The circadian clock also plays a broader role in maintaining health during aging because mice deficient in *BMAL1* in all tissues have reduced lifespan and several symptoms of premature aging including cataracts and neurodegeneration (132–134). Moreover, homozygous *Per^{luc}* mice exhibit age-dependent

photoreceptor degeneration and premature aging of the retinal pigment epithelium (135). Since the PER::LUCIFERASE fusion protein is wild type and is used as a model for studying circadian rhythms (136), this observation suggests that even very mild changes in the expression or function of core circadian clock regulators can negatively impact the health of the aging eye. Notably, disruption of circadian rhythms in the human eye contributes to glaucoma and is also implicated in development of myopia (137,138). We note that while preparing this manuscript, a preprint showed that expression of Clk^{DN} under the photoreceptor-specific trpl-Gal4 driver caused decreased phototactic behavior in young flies relative to an age-matched control (139), which is consistent with the rhabdomere loss observed in our study. However, rhabdomere integrity was not tested by the authors of this study. Therefore, our studies suggest that the *Drosophila* retina serves as a useful model to study circadian-dependent regulation of photoreceptor homeostasis. It is important to note that Clk and Cyc could also have Clock-independent roles, as shown previously for aging brains (120). Therefore, it will be important for future studies to establish Clock-dependent and -independent functions of Clk and Cyc in fly photoreceptors.

How does disruption of the circadian clock in photoreceptors lead to light-dependent retinal degeneration? Our data suggest that although Clk-dependent transcription is necessary to prevent high levels of oxidative stress in the eye, this is not the proximal cause of retinal degeneration in young flies. Instead, we favor a model in which Clk:Cyc directly binds and activates expression of genes encoding the phototransduction machinery in photoreceptors, maintaining the continued expression of these proteins that have an integral role in photoreceptor structure and function. Numerous studies have demonstrated that complete loss of function of individual phototransduction genes results in retinal degeneration, often dependent on light exposure (118). Our data show a significant and substantial decrease in transcript levels of multiple phototransduction genes, and we propose that the cumulative decrease in expression of their corresponding proteins causes the light-dependent retinal degeneration in flies expressing Clk^{DN}. However, because disrupting Clk:Cyc activity in photoreceptors had widespread effects on gene expression and chromatin accessibility, we cannot exclude the possibility that other pathways such as autophagy (140), Ca^{2+} signaling (141) and phosphoinositide metabolism (142) also contribute to the onset of light-dependent retinal degeneration.

The circadian clock has been implicated in the oxidative stress response in *Drosophila* (120,143) and in mammals (144). Since the onset of many age-related eye diseases is particularly sensitive to disruptions in oxidative stress response (145), an increase in ROS levels in Rh1>Clk^{DN} eyes suggests an overall neuroprotective role for the circadian clock regulators in the retina, one which might become increasingly important with advanced age. Under standard laboratory conditions, wild-type flies with pigmented eyes only begin to exhibit the first signs of retinal degeneration after D50 (146). Thus, we hypothesize that the increase in Clk:Cyc activity in aging photoreceptors protects against retinal degeneration in part, by promoting expression of genes that combat oxidative stress. Supporting this hypothesis, several stress response genes exhibit cyclic expression patterns in the head only in older flies (D55), and this cyclic expression is dependent on both Clk and oxidative stress (126). However, because these studies were performed in female white-eyed w^{1118} flies (126), which already show substantial retinal degeneration by D15 to D30 (147), it remains to be elucidated if these age-associated changes in cyclic transcription also occur in flies with pigmented eyes.

Here, we identified an age-dependent increase in Clk and Cyc TF activity in photoreceptors via an unbiased integrative ATAC-seq and RNA-seq approach, which focused on identifying changes in TF activity during aging based on changes in chromatin accessibility around TF binding sites. Thus, our initial approach was not focused on circadian biology and the data was obtained from single time-point comparisons; samples were collected at ZT6 +/- 1 for aging experiments and at ZT9 +/- 1 for the Clk^{DN} experiments. Because of these sampling differences, the increase in Clk:Cyc TF activity in aging could reflect a change in the phase of Clk:Cyc binding to DNA during the day, as shown for Monarch butterfly brains (148), an increase in amplitude of binding activity, or both. Supporting the latter possibility, circadian analysis of the transcriptome of aging fly heads showed both an increase in the amplitude and a slight shift in phase of *tim* and *per* expression, moving earlier in the day as flies aged (126). Other studies have observed a decrease in protein levels of the Clock repressor PER with age in *Drosophila* photoreceptors but not in pacemaker neurons (149,150), suggesting that the increased Clk:Cyc activity observed in our study could in fact reflect an increase in DNA binding of Clk:Cyc in old flies mediated by decreased repression by the Per:Tim complex. Nonetheless, the mechanisms underlying the age-associated changes in the circadian clock and Clk:Cyc TF activity in fly photoreceptors remain to be elucidated.

Overall, our work identifies a central role for the circadian clock in regulating the photoreceptor transcriptome. We observed that Clk:Cyc contributes to the expression of many thousands of genes in adult photoreceptors, consistent with reports from mammalian cells showing that Clock activity regulates the expression of 15% of expressed genes (116). Clk and Cyc have been proposed to act at the top of a regulatory hierarchy to control widespread cyclic transcription in many cell types by regulating the expression of TFs (21,22). Our data from photoreceptors are consistent with Clk:Cyc regulating expression of several important eye-specific TFs in *Drosophila* photoreceptors, suggesting a mechanism through which Clk and Cyc control expression of many genes indirectly. This is likely a widespread phenomenon across *Drosophila* tissues because gene regulatory network analysis of cyclic transcripts in brain, gut, Malphigian tubules, and fat bodies also identified many Clock-regulated TFs, including *h, hth, Mitf, and GATAd* (151).

Interestingly, *per* (*per^{01}*) and *tim* (*tim^{01}*) male mutants have extended lifespan, mediated by a loss of PER in intestine cells (152), suggesting that increased Clock activity in a particular tissue can lead to positive outcomes associated with health- and lifespan in a tissue- and sex-specific manner. Thus, identifying the molecular mechanism that underlie the age-associated changes in Clk:Cyc activity in aging photoreceptors and their impact on cellular homeostasis, which may be specific to the retina or other peripheral clocks in flies, will be an important area of research for future studies.

2.4   Methods

2.4.1   Fly collection and maintenance

Rh1-Gal4>UAS-GFP^{KASH} (*w^{1118};; P{w^{+mC}=[UAS-GFP-Msp300KASH}attP2, P{ry$^{+t7.2}$=rh1-GAL4}3, ry^{506}*) (13) flies were maintained on standard fly food as previously described. For aging experiments, flies were collected in a 3 day window after eclosion and transferred to population cages. For Clk^{DN} experiments, flies were collected in a 24 hour period and transferred to population cages. D1 corresponds to flies that were collected the first day immediately after eclosion. Flies for both aging and Clk^{DN} experiments were maintained in a 25°C incubator with a 12:12 hour light:dark cycle, where relative Zeitgeber Time (ZT) 0 corresponds to when the light cycle begins. For aging experiments, male flies were collected and flash-frozen at ZT6 (-/+) 1 hour. For Clk^{DN} experiments, male flies were collected and flash-frozen at ZT9 (-/+) 1 hour. We note that UAS-

GFP[KASH] and additional tagged KASH or QUAS flies (87) are available at Bloomington Drosophila Stock Center (BDSC) for nuclei immunoprecipitation in different tissues (#92580 for the UAS-GFP[KASH] flies used in this study). Rh1-Gal4 (BDSC #8691), UAS-Clk[DN] [#1] (BDSC #36319), UAS-Clk[DN] [#2] (BDSC #36318) and UAS-LacZ (BDSC #8529) fly lines were obtained from BDSC. UAS-Cyc[DN] were generously provided by Dr. Daniel Cavanaugh (Loyola University)

2.4.2 Chromatin accessibility and transcriptome analysis of photoreceptors

Nuclei immuno-enrichment (NIE), Omni-ATAC, quantitative PCR (qPCR), and RNA-seq were performed on male *Rh1-Gal4>UAS-GFP[KASH]* flies at the indicated ages as previously described (87). Briefly, for NIE experiments, we processed 400 age-matched male flies per replicate. Three independent biological replicates were obtained and analyzed for all RNA-seq and ATAC-seq samples. RNA-seq and ATAC-seq data analysis were performed as described in (87,146); details specific to this study are described below.

Primers for quantitative PCR

Primer name	Sequence (5' – 3')
ClkDN_Fwd	CGACAAGGATGATACAGATCAG
ClkWT_Fwd	GCGAGAAGAAGCGACGAGAT
Clk_Rev	ATTGCTGAGGAACGCAGGCT
eIF1a_Fwd	GCTGGGCAACGGTCGTCTGGAGGC
eIF1a_Rev	CGTCTTCAGGTTCCTGGCCTCGTCCGG
Rpl32_Fwd	GCTAAGCTGTCGCACAAATG
Rpl32_Rev	CGTTGTGCACCAGGAACTT

Note: We used the same Clk_Rev (reverse) primer to detect either WT or DN transcript by combining with corresponding Forward primer (Fwd).

RNA-seq fold change heatmap analysis: Heatmaps were obtained using $\log_2$ transformed fold change values obtained from pair-wise comparisons using DESeq2 (153). For aging heatmaps, DEGs obtained from each comparison with D10 were pooled and de-duplicated. Plots were obtained using R (v4.0.3) package *pheatmap* (v1.0.12).

Differential TF activity using diffTF: DiffTF (92) analysis was performed using the default parameters: analytical approach (*nPermutations=0)*, integration mode using raw counts obtained from STAR *(--quantMode GeneCounts)* (65), narrow peaks obtained using MACS2 (71). TFBS bed files were obtained from PWMScan (154) using the Aug 2014 BDGP Rel6 + ISO1 MT/dm6 target genome with the 353 available motifs from the CIS-BP library (94) with default parameters *(p-value<0.00001, Bg base composition 0.29,0.21,0.21,0.29).* Identified TFs were classified as significant if they had an FDR lower than 0.001 in at least *one* of the pair-wise aging comparisons with D10, or in either D1 or D10 analysis for the Clk^{DN} comparisons. Detailed protocols for NIE, RNA-seq, Omni-ATAC, ChIP-seq, and CUT&Tag are available at: dx.doi.org/10.17504/protocols.io.buiqnudw.

Clk ChIP-seq peak and photoreceptor ATAC-seq overlap: Previously published high-confidence peaks obtained from Clk and Cyc ChIP-seq data (114) were downloaded, and genomic coordinates were transformed from the *dm3* to *dm6* genome using the *LiftOver* tool from the UCSC Genome Browser website (155). Genomic overlap was calculated using R (v4.0.3) packages *ChIPpeakAnno* (v3.24.2) and *GenomicRanges* (v.1.42.0).

Transcript per million (TPM) scatter plots: TPM values for each sample and replicate were obtained using TPMCalculator (156). TPMs were averaged between biological replicates and used for scatter plot generation using R package *ggplot2* (v3.3.3)*.* DEGs obtained with DESeq2 were colored on the TPM scatter plots.

2.4.3   Optic neutralization

Optic neutralization and retinal degeneration quantification were performed as previously described (10,157). Briefly, flies were glued to a glass slide using transparent nail-polish, and eyes imaged using bright-field microscopy. We note that optical neutralization using bright-field microscopy (with white light) is only possible with flies that have pigmented eyes, which is the case for all the flies tested in the present manuscript. Five biological replicates were analyzed for each sample, and retinal degeneration scores were assessed blindly.

2.4.4 Targeted GSH:GSSG metabolomic assay

25 eyes per sample were collected from male flies of the indicated age, condition, and genotype (n = 3). Eyes were placed in a Covaris MicroTube with 110 µl of blocking solution (62.5 mg NEM, 3 mg EDTA, 5 mg $NaHO_3$ disolved in 2 mL of 3:2 parts Water/MeOH (v/v)). Once all eyes were added to the blocking solution 10 µl of 100ng/ul of each internal standard was added $^{13}C_2$-^{15}N-GSH-NEM and $^{13}C_4$-$^{15}N_2$-GSSG. Samples were homogenized using a Covaris Ultra Sonicator with the following settings: peak power: 200; duty factor: 10%; cycles per burst: 200; time: 300 seconds. Samples were then processed using an Agilent 1260 Rapid Resolution liquid chromatography (LC) system coupled to an Agilent 6470 series QQQ mass spectrometer (MS/MS) to analyze glutathione (158). (Agilent Technologies, Santa Clara, CA). A Waters T3 2.1 mm x 100 mm, 1.7 µm column was used for LC separation (Water Corp, Milford, MA). The buffers were A) water + 0.1% formic acid and B) acetonitrile + 0.1% formic acid. The linear LC gradient was as follows: time 0 minutes, 0 % B; time 2 minutes, 0 % B; time 8 minutes, 30 % B; time 9 minutes, 95 % B; time 9.1 minutes, 0 % B; time 15 minutes, 0 % B. The flow rate was 0.3 mL/min. Multiple reaction monitoring was used for MS analysis. Data were acquired in positive electrospray ionization (ESI) mode. The jet stream ESI interface had a gas temperature of 350°C, gas flow rate of 9 L/minute, nebulizer pressure of 35 psi, sheath gas temperature of 300°C, sheath gas flow rate of 9 L/minute, capillary voltage of 4000 V in positive mode, and nozzle voltage of 1000 V. The ΔEMV voltage was 450 V. Agilent Masshunter Quantitative analysis software was used for data analysis (version 8.0).
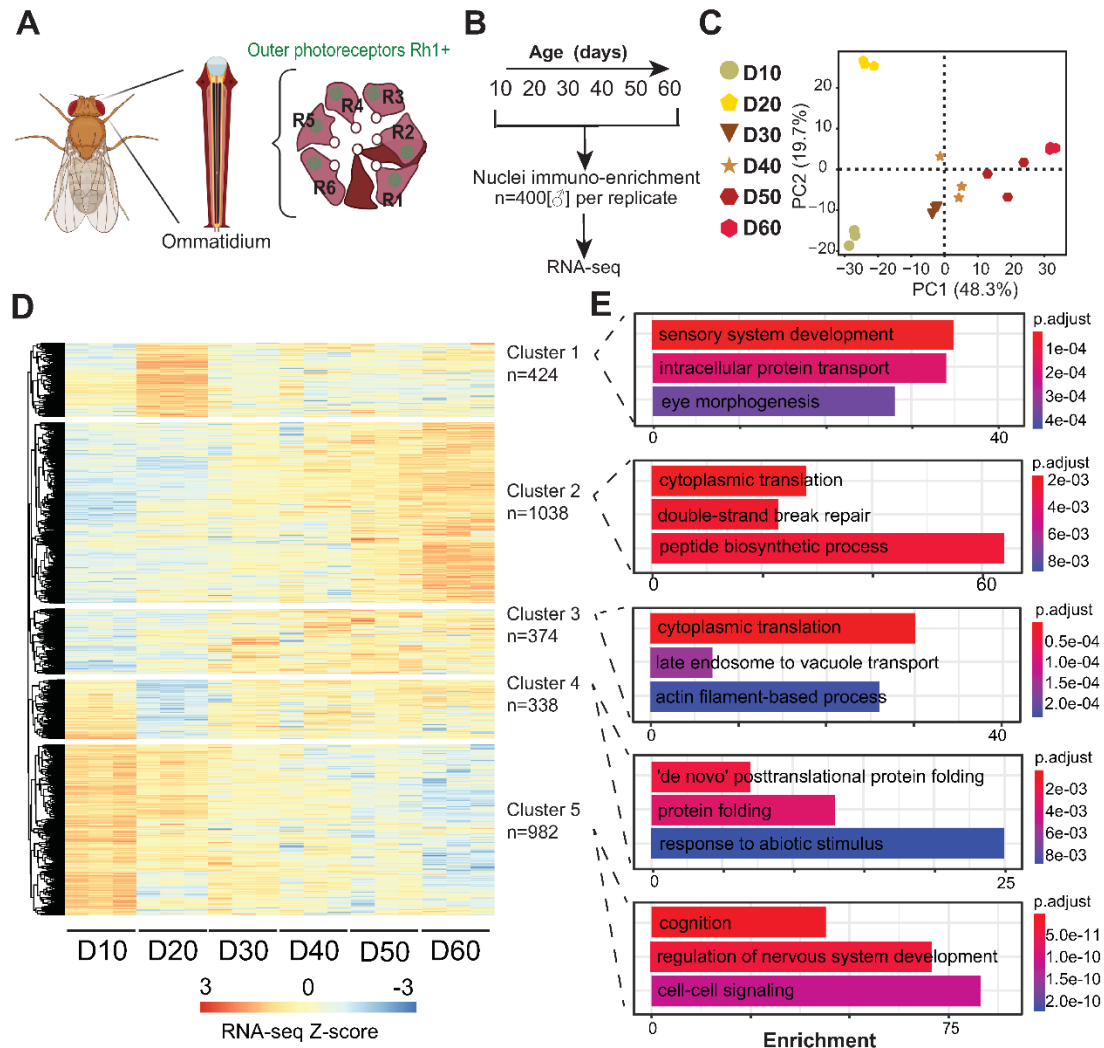
2.5 Figures

Figure 2.1 A. Schematic of the *Drosophila* compound eye, composed of ommatidia that contain six outer photoreceptors that express Rhodopsin 1 (Rh1). Outer photoreceptor nuclei are labeled with GFP[KASH]. B. Experimental design for the aging RNA-seq time course. C. Principal Component Analysis (PCA) of aging RNA-seq samples based on gene counts. D. Hierarchically clustered heatmap of aging RNA-seq samples. Only genes that were identified as being differentially expressed in any condition relative to D10 are shown. Z-scores are calculated based on normalized counts obtained using DESeq2, and the heatmap is divided into five clusters based on the dendrogram. E. Gene Ontology (GO) term analysis of significantly enriched genes in each cluster.
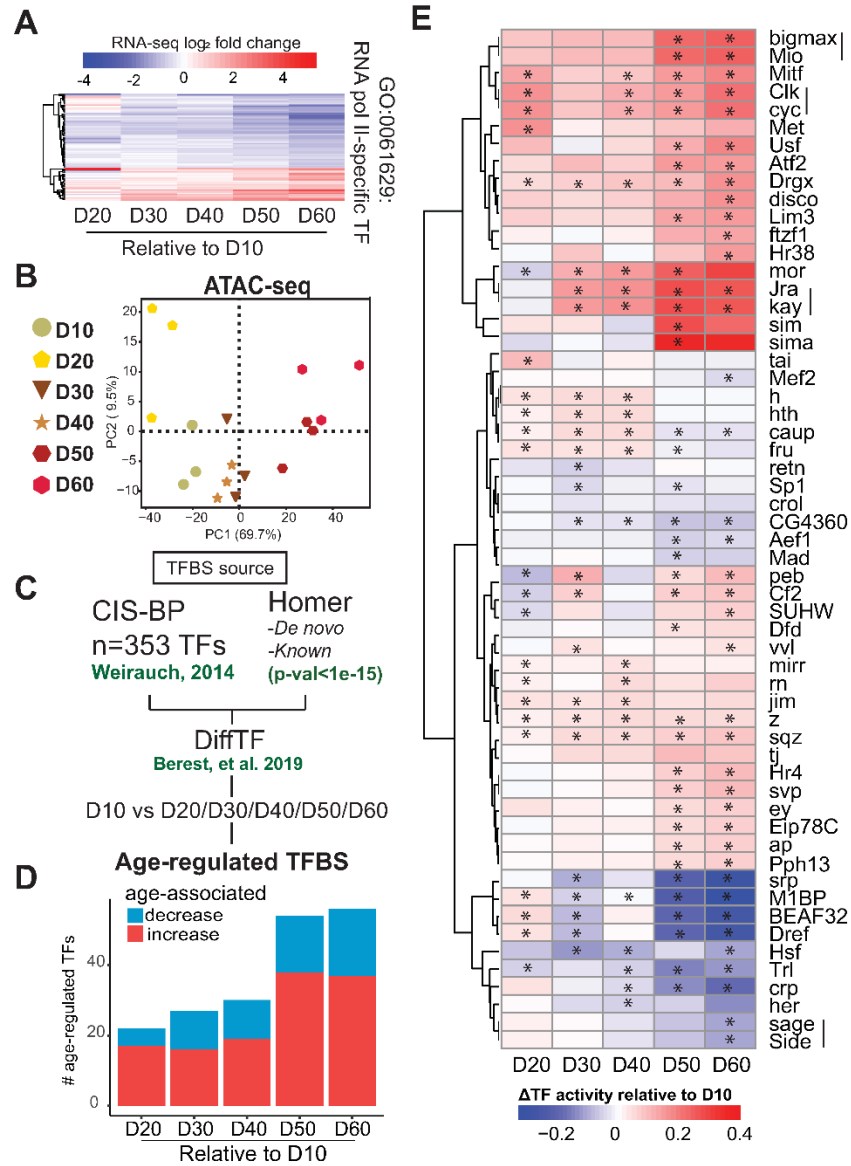
Figure 2.2 A. Heatmap of RNA-seq fold change during aging for significantly differentially expressed genes (DEGs; p-adj<0.05, |FC|>1.5 relative to D10) associated with the GO term "RNA polymerase II-specific DNA-binding transcription factor binding" (GO:0061629). Genes are hierarchically clustered by $\log_2$ fold change values. B. PCA of aging ATAC-seq samples based on the read distribution over binned genome. C. Schematic of differential transcription factor (TF) analysis approach using *diffTF*. D. Bar plot indicating the number of significant age-regulated TFs at each age relative to D10 (p-adj<0.001). E. Hierarchical clustered heatmap of age-regulated TFs with significant changes in activity between any age and D10 (asterisk, p-adj<0.01). Scale represents the relative change in TF activity with red showing higher TF activity in old samples relative to D10, and blue indicating an age-associated decrease in activity. TFs that bind a common motif and cluster together are indicated by a vertical line.
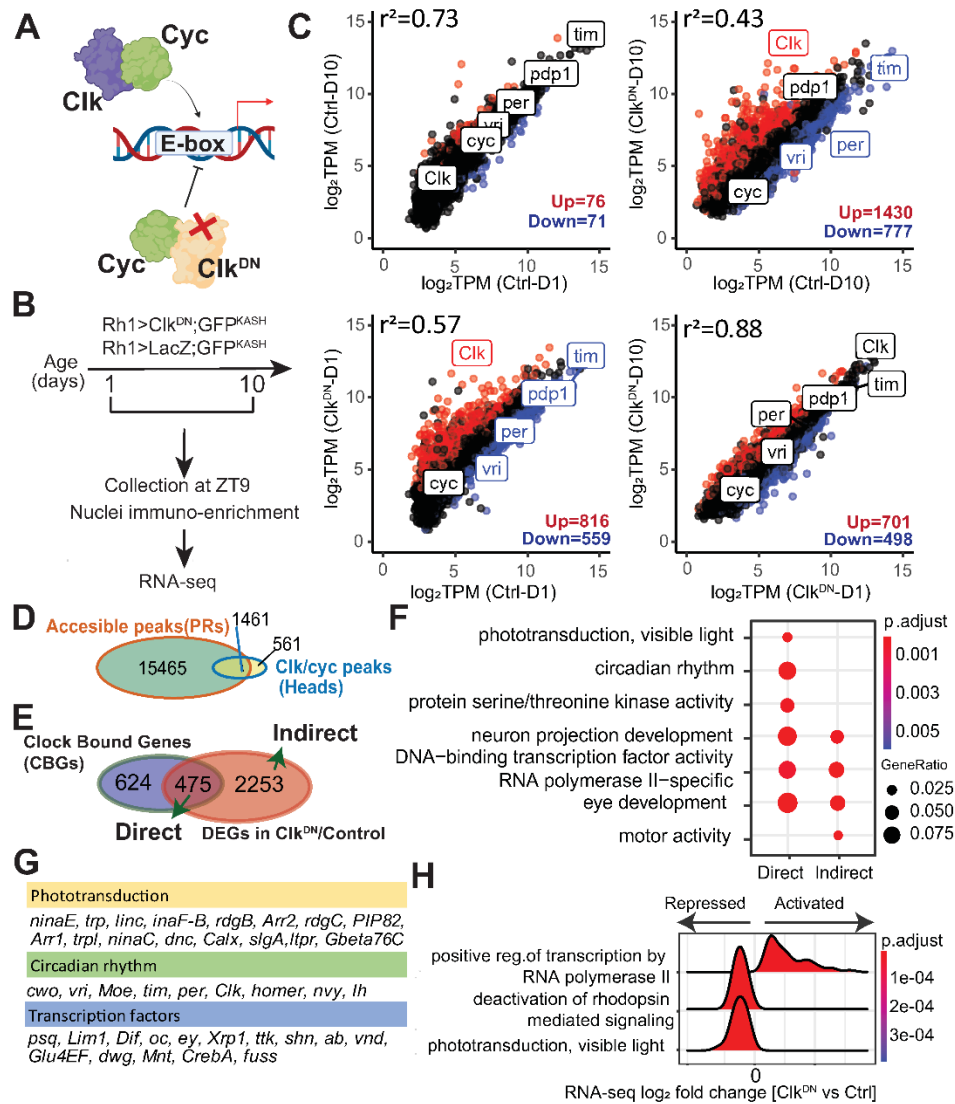
Figure 2.3 A. Schematic of Clk:Cyc activity. Expression of Clk[DN] disrupts Clock-dependent transcription because it binds Cyc but does not contain the DNA binding domain. B. Schematic for the RNA-seq experiments. Flies that express the GFP[KASH] tag and either Clk[DN] or LacZ (control) in photoreceptors were aged to D1 or D10 prior to NIE. Flies were harvested at the indicated ZT. C. Scatter plot of mean expression (TPM, n=3) for the indicated pair-wise comparisons. Up- and down-regulated DEGs (p-adj<0.05, |FC|>1.5) are colored in red and blue, respectively. Core clock genes *Clk*, *cyc*, *vri*, *per*, *Pdp1* and *tim* are highlighted. D. Genomic overlap of accessible peaks in photoreceptors and Clk:Cyc binding sites identified using ChIP-seq in (114). E. Venn diagram comparing overlap of Clock-bound genes (CBGs-green) with genes that were differentially expressed (adj p <0.05, |FC|>1.5) in Rh1>Clk[DN] relative to control. F. GO term analysis of Clock direct and indirect targets. G. Table showing selected genes identified as Clock direct target genes that were associated with the indicated biological processes. H. Ridge plot of selected GO terms in flies expressing Clk[DN] versus control at D1 identified using Gene Set Enrichment Analysis.
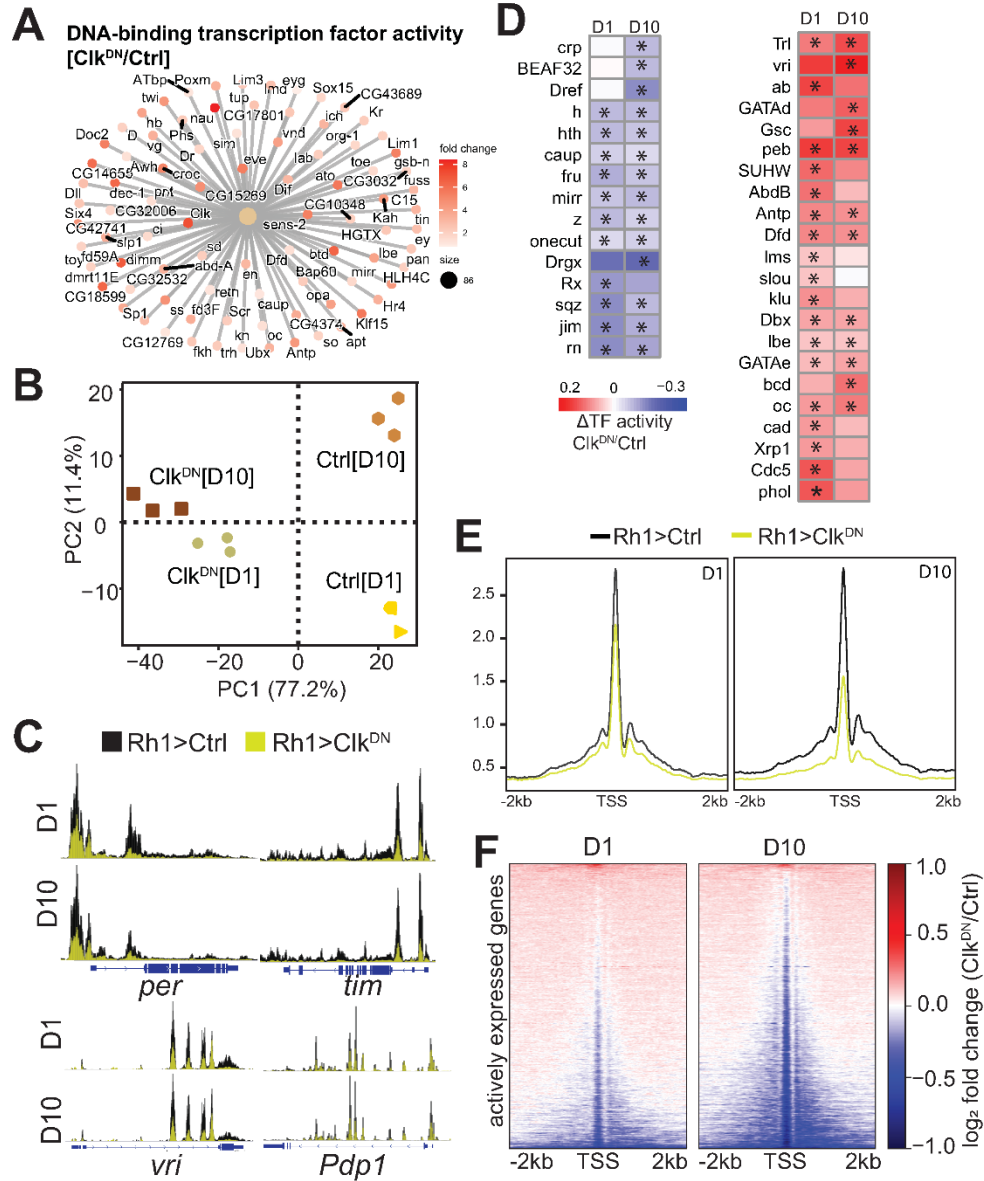
Figure 2.4 A. Cnet plot of genes associated with the GO term "DNA-binding transcription factor activity". Color represents change in expression in Clk^{DN} relative to control at D1. B. PCA of Clk^{DN} and control ATAC-seq samples based on the read distribution over a 10-kb binned genome. C. Genome browser inspection of CPM-normalized ATAC-seq signal for selected core Clock genes (*per*, *tim*, *vri* and *Pdp1*). Scale is normalized to the same height in each comparison. ATAC-seq from Ctrl flies is labeled in black, and Clk^{DN} in yellow. D. Hierarchical clustered heatmap of significant Clock-regulated transcription factors identified between Rh1>Clk^{DN} and Rh1>Ctrl at either D1 or D10. Scale represents the relative change in TF activity. E. CPM-normalized gene metaplots of ATAC-seq signal centered around Transcription Start Sites (TSS). F. ATAC-seq fold change heatmaps of signal centered around TSS for all actively expressed genes in photoreceptors. Fold change is obtained by dividing signal from Rh1>Clk^{DN} relative to Rh1>Ctrl, and log^2 transformed to center changes around zero (no change).
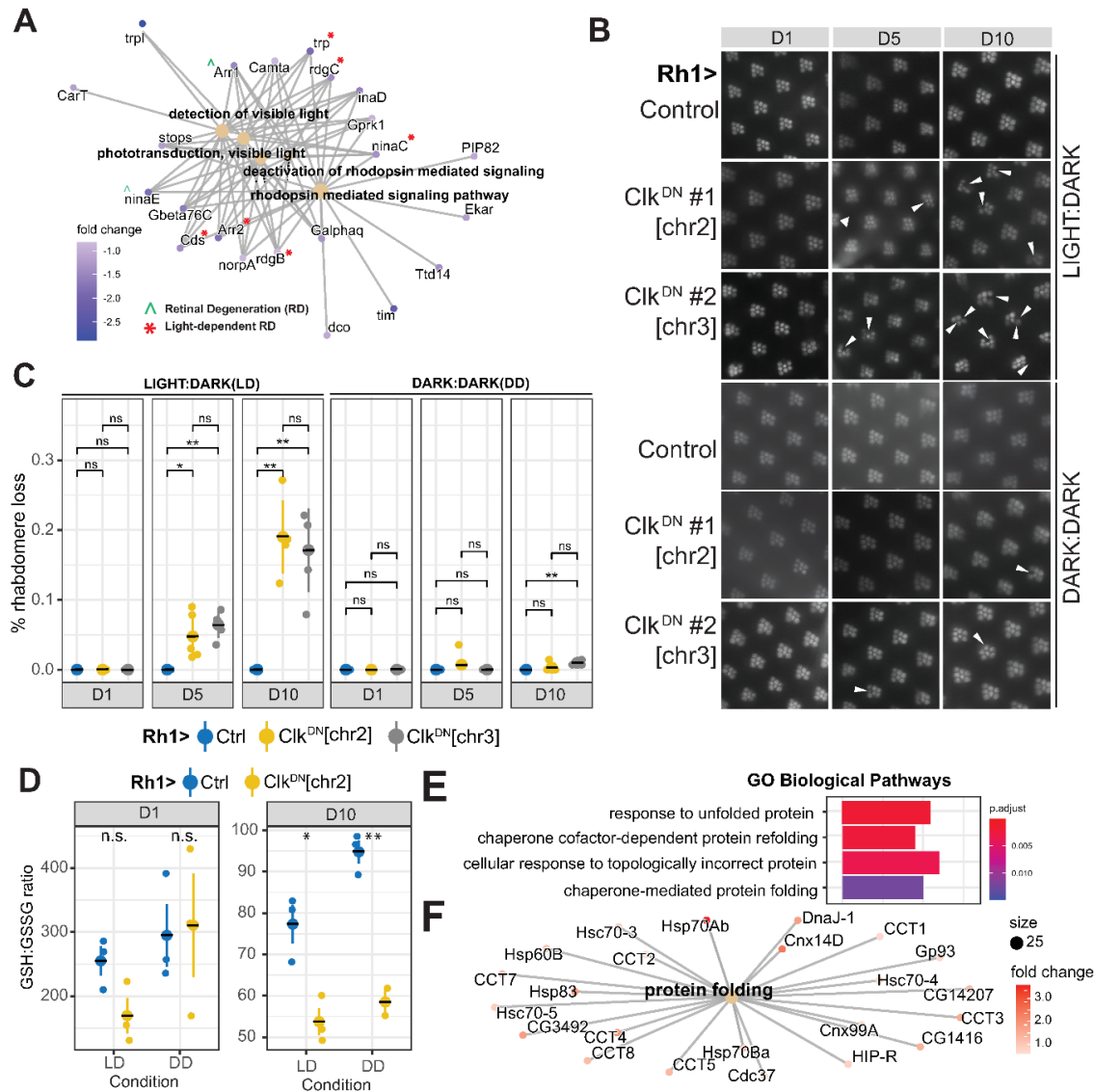
Figure 2.5 A. Cnet plot of DEGs (p-adj<0.05,|FC|>1.5) in Rh1>Clk^{DN} relative to Rh1>Ctrl at D1, associated with phototransduction-related GO terms. Genes associated with light-dependent and - independent retinal degeneration based on published literature are indicated by accent or asterisks. B. Representative images of eyes from flies expressing Clk^{DN} versus control at the indicated age reared in light:dark (LD – top) or constant dark (DD - bottom) conditions. Images were obtained using optical neutralization (see Methods). Arrows indicate position of missing/degenerated rhabdomeres. C. Scatter plot showing quantification of retinal degeneration in panel B. Individual eyes are shown as circles with means shown as horizontal black lines overlaying large circle (n=5). The distribution for each LD reared sample was compared with the DD control for the same age using ANOVA followed by Tukey-HSD comparing ages, samples, and condition (L:D vs D:D). "n.s." = not significant, "*" = FDR<0.05, "**" = FDR<0.005). D. Reduced(GSH):Oxidized(GSSG) Glutathione ratios in dissected eyes from flies of the indicated genotype, age, and condition. Statistics were performed as in panel C. E. Bar plots representing enriched GO terms amongst the genes that were upregulated in Rh1>Clk^{DN} at D10 relative to D1. F. Cnet plots of genes identified in panel E.

CHAPTER 3.   PROPER CONTROL OF R-LOOP HOMEOSTASIS IS REQUIRED FOR MAINTENANCE OF GENE EXPRESSION AND NEURONAL FUNCTION DURING AGING

Declaration of collaborative work

The work described in this chapter was the collaborative work of Juan Jauregui-Lozano, Spencer Escobedo, Alyssa Easton, Nadia Atallah, and Vikki Weake, under the supervision of Hana Hall. Spencer Escobedo performed the phototaxis assays. Alyssa Easton and Hana Hall performed the slot blots. Nadia Atallah and Vikki Weake assisted with data analysis. All other experiments, as well as data visualization were performed by Juan Jauregui-Lozano. The original draft was written and edited by Juan Jauregui-Lozano and Hana Hall. This work was a result from a collaboration with Dr. Hana Hall.

3.1   Introduction

Aging is a process characterized by a time-dependent decline in physiological homeostasis that eventually leads to a loss of organismal function and increased incidence of death. Characteristic functional changes include loss of gene and protein expression, mitochondrial dysfunction, cellular senescence, and stem cell exhaustion (1). Aging is also a major contributor to development of many chronic diseases including ocular disease (6). Age-related vision loss and maculopathy have been associated with decreased density of retinal cells, including photoreceptors. Specifically, the age-related decline in photoreceptors affects predominantly rods rather than cones (159). In addition, emerging evidence links age-related neurological diseases, including retinal neuropathies, to defects in gene expression and RNA metabolism (75). Nonetheless, the molecular mechanisms that contribute to the age-associated susceptibility of the eye to disease development are poorly understood.

R-loops are three-stranded nucleic acid structures consisting of an RNA-DNA hybrid and a misplaced single strand of DNA (160). They typically form during transcription in organisms ranging from yeast to humans and play a significant role in normal cellular physiology, being required for the initiation of mitochondrial replication and class switch recombination. Notably, recent studies suggest that R-loops can dynamically regulate gene expression (161); for example,

due to their enrichment at gene termini, R-loops can modulate gene expression by preventing DNA methylation or limiting transcription factor access to promoters and facilitating efficient transcription termination at 3'-ends (Boque-Sastre et al. 2015; Skourti-Stathaki et al. 2014). Moreover, R-loop formation correlates positively with active transcription, gene length, GC content and DNA topology (165). Importantly, recent studies have shown that resolution of topological stress during transcription mediated by topoisomerases is critical for proper neuronal function (166). Topoisomerase 3β (Top3β) is a member of the type IA superfamily of topoisomerases, which unwind negatively supercoiled DNA formed during transcription and replication, an activity that prevents R-loop formation (165). Loss of Top3β is associated with neurological disorders (165) and has been shown to reduce lifespan in mice (167).

Although R-loops are normal biological structures, their persistent formation is a major source of spontaneous DNA damage that can lead to transcriptional dysregulation and genome instability (160); two early hallmarks of aging. Given that aging is a main risk factor for development of chronic diseases, including neurodegenerative disease, it is conceivable that R-loops could play a significant role in age-associated mis-regulation of cellular functions. However, our understanding on R-loop biology during aging, and particularly in neurons, is quite limited. Due to high levels of transcription and alternative splicing, retinal cells and particularly photoreceptor neurons may be highly sensitive to RNA metabolism dysregulation. *Drosophila* compound eye contains approximately 800 units called ommatidia with each consisting of 20 cells, including eight photoreceptor (PR) neurons. The six outer PRs (R1-R6) expressing Rhodopsin 1 (Rh1) are mainly responsible for black and white vision and motion, and are similar to human rods. The inner PRs (R7 and R8) express Rh3/4 and Rh5/6, respectively, are responsible for color vision and resemble human cones. To characterize how aging impacts the genomic R-loop landscape in *Drosophila* photoreceptors, we isolated genetically labeled outer PRs using our recently improved nuclei immuno-enrichment method (5) and performed MapR coupled with next generation sequencing (168). Here, we show that R-loop levels in photoreceptor neurons increased progressively with age and were associated with genic characteristics, such as transcript levels and GC content. Further, our data show that majority of genes that decrease expression during aging contained R-loops. Finally, photoreceptor-specific depletion of DNA/RNA topoisomerase Top3β resulted in increased R-loop levels, reduced expression of long genes with neuronal function and reduced visual response. Importantly, overexpression of either Top3β or human RNASEH1, an enzyme that

resolves R-loops, in the eyes resulted in enhanced visual response during aging. Together, our data show that aging is associated with increased levels of R-loops over transcribed genes, potentially disrupting transcriptional outcomes that contribute to age-associated changes in neuronal function, including visual response to light.

3.2 Results

3.2.1 Aging photoreceptor neurons show increased global R-loop levels that correlate with loss of function and precede age-associated retinal degeneration

To examine the global R-loop levels in photoreceptor neurons, we tagged the outer nuclear membrane of R1-R6 with GFP fused to the KASH domain of Msp300 protein using the Rh1-Gal4 driver and isolated outer PR nuclei from the head homogenate with our tissue-specific method (5). Next, we aged flies over the course of 50 days post-eclosion (emergence from the pupae) and extracted genomic DNA from isolated PR nuclei at three time points (Figure 1A). We then assessed the global R-loop signal with a DNA slot blot assay using the S9.6 antibody, which recognizes RNA-DNA hybrids. Specificity of S9.6 antibody towards RNA-DNA hybrids was shown by pre-treatment of DNA with ribonuclease H1 (RNase H1) that resulted in significant decrease of S9.6 signal (Figure 1B). Signal quantification showed approximately 30% increase in R-loop levels in PRs isolated from middle-aged, 30-day old flies as compared to that in young, 10-day old flies. This trend continued with a significant increase of nearly 50% in global R-loop levels at day 50 (Figure 1B-C). These data show that R-loops start accumulating early during photoreceptor aging, at a time point where flies show decreased visual function (170). Importantly, using optical neutralization, which measures photoreceptor structural integrity with light microscopy, we observed no retinal degeneration by middle age, with a stochastic loss of rhabdomeres occurring after day 40. Thus, our data show that process of R-loop accumulation precedes age-related retinal degeneration, thus suggesting that increased R-loop formation may contribute to loss of function and possibly neuronal cell loss during aging.

3.2.2 Profiling genome-wide R-loop distribution in PR neurons reveals age-associated changes

To determine the genomic R-loop landscape in aging PRs, we coupled our NIE approach with MapR, a recently published R-loop mapping strategy based on the specificity of RNase H1 enzyme

to RNA-DNA hybrids, combined with the micrococcal nuclease (MNase)-based CUT&RUN technology. MapR uses a recombinant mutant form of MNase-fused RNase H1, which binds but does not degrade the RNA moiety within an RNA-DNA hybrid (ΔRH). Upon binding of the RNA-DNA hybrid by ΔRH-MNase, MNase activation by Ca^{2+} addition results in cleavage of surrounding DNA and a subsequent release of R-loop associated DNA fragment, which is used for library preparation coupled with high-throughput sequencing (Figure 2A). Surprisingly, we found that coupling NIE-purified photoreceptor nuclei with the standard MapR protocol yielded signal over genic regions resembling MNase-seq rather than R-loop specific enrichment. Our data showed MapR signal depletion around the Transcription Start Site (TSS) of genic regions, suggesting that our samples were being over-digested by MNase. We therefore modified the standard MapR protocol based on the recently published improvement of the CUT&RUN method, which incorporates high salt with low calcium washing steps, and also decreased digestion time (see Methods in Supporting Information). To evaluate the quality of our modified protocol, we first compared our MapR data in *Drosophila* PRs to the original MapR data obtained in human HEK293T cells and found that the metagene profiles over the gene bodies showed similar R-loop distribution, with signal enrichment around the TSS and the Transcription Termination Site (TTS) . In contrast, when we next compared our MapR data with R-loop mapping data obtained from *Drosophila* embryos using DRIP-seq (171), we observed that DRIP-seq showed signal enrichment over gene bodies with a slight depletion around TSS; discrepancies that have been shown previously to result from different affinities of S9.6 antibody used in DRIP-seq, as opposed to RNase H1 used in MapR, for RNA-DNA hybrids. In addition, the Alecki dataset was obtained during *Drosophila* early embryonic developmental stages. Thus, some differences in signal enrichment in our studies might arise from tissue-specific and method-specific effects.

To validate the specificity of our MapR method, we treated NIE-purified PR nuclei with RNase-H1 prior to performing MapR, which led to a complete loss of signal enrichment as shown by individual gene examples (Figure 2B) and gene metaplots (Figure 2C). In addition, when we normalized MapR signal by obtaining a ratio from RNaseH1 non-treated relative to treated samples, we found that the MapR signal distribution did not show significant changes in enrichment over genes bodies. We note that the standard MapR protocol includes a separate negative control where nuclei are incubated with MNase alone to account for MNase binding.

However, in our hands, the MNase control yielded no purifiable DNA, as shown by TapeStation profiles.

Based on these observations, we performed MapR in aging PRs at day 10, 30 and 50 post-eclosion using our modified approach in three independent biological replicates which generated at least 3.5×10^7 uniquely mapped fragments per sample. Spearman's correlation analysis based on read distribution over a 1000-bp binned genome revealed a strong positive association amongst the three biological replicates (Spearman's $p \geq 0.96$). In contrast, when we compared the samples between different age time points, we observed lower positive association between day 10 and day 50 (Spearman's $p \geq 0.93$), suggesting that the R-loop landscape changes were age-dependent (Figure 2D). Similarly, Principal Component Analysis (PCA) of the normalized R-loop distribution revealed that 53.8% variance amongst the samples for all biological replicates was attributed to age (Figure 2E). Notably, while samples clustered by age, the similarity amongst the biological replicates decreased with age (PC2 13.6%), suggesting that aging is associated with increased heterogeneity in R-loop distribution.

To further assess the quality of our sequencing datasets, we performed peak calling using the Model-based Analysis for ChIP-Seq (MACS2) algorithm using default settings and measured quality control metrics. To account for the differences in the number of mapped fragments in each sample, we called the peaks using bam files that were down-sampled to the same number of mapped fragments (3.5×10^6). Evaluation of the Fraction of Reads in Peaks (FRiP) score, which measures the quality of signal enrichment as defined by modENCODE, showed consisted FRiP scores higher than 0.37 for all samples. Furthermore, we found that peak distribution was stably maintained during aging, with approximately 60% of peaks being annotated to promoters (TSS ± 2kb) and approximately 25% of peaks annotated to introns, which is consistent with previously reported genome-wide R-loop distribution.

Taken together, our data demonstrate that we successfully applied the MapR method to tissue-specific samples in *Drosophila,* by purifying photoreceptor nuclei from the whole organism, and produced high quality R-loop mapping data. Furthermore, application of MapR in aging PRs showed that the genome-wide R-loop distribution changes in an age-dependent manner.

3.2.3 Age-associated R-loop accumulation over gene bodies correlates with high GC content, gene length and transcriptional levels

Given that R-loops typically form co-transcriptionally (172), we next examined global and locus-specific distribution of age-associated changes in R-loops across actively transcribed genes, defined as having more than seven transcripts per million (TPM). First, we analyzed global R-loop signal over gene bodies for actively expressed genes and compared the average signal across the gene, as counts per million (CPM). As expected, gene metaplots revealed that R-loop signal was enriched mainly over TSS and towards the 3' ends of genes across all age time points (Figure 3A), which is consistent with previously reported R-loop distribution (Ginno et al. 2012; Sanz et al. 2016a; Nadel et al. 2015; Chen et al. 2017). Importantly, there was a widespread increase in R-loop levels over gene bodies during aging, most notably at day 50. Genome browser inspection of the averaged signal tracks for each time point for two individual genes showed changes in R-loop signal, including an early decrease or late increase. To further evaluate the molecular characteristics of R-loop coverage during aging, we asked whether the age-associated increase in R-loops could be a consequence of broadening of the peaks. Consistent with our gene metaplots, MapR signal around the peaks showed that overall, R-loop coverage increased with age (Figure 3B), suggesting that R-loops might either extend or form at higher rate with age. To further assess an age-associated increase in R-loop occupancy, we quantified signal coverage, as defined by the sum of peak width for each time point. Notably, R-loop peaks covered approximately 18.7 megabases (Mb) of the genome at day 50, compared to 18.1 Mb at day 10, showing a modest but significant increase in coverage during aging (t-test, $p<0.022$) (Figure 3C). Supporting this data, violin plots depicting the peak width for all peaks revealed a slight but consistent increase in peak width at day 30 and day 50 as compared to that at day 10. Taken together, these observations show that R-loop signal over the genome accumulates with age, corroborating our findings from bulk R-loop levels using slot blots (see Figure 1). We note that the *Drosophila* genome has a total size of 180 Mb, and our data in *Drosophila* PRs showed that R-loops covered approximately 10% of the genome, which is similar to the genomic R-loop coverage obtained from other organisms, including mammals (160).

To gain further insight into age-associated changes in R-loops, we compared the fold change in R-loop signal around TSS (±3 kb) or TTS (±3 kb) in old and young PRs. Heatmap plots of actively expressed genes ranked based on their fold change showed that the majority of the TSS-associated

R-loops increased with age, while only approximately 50% of genes had an increase in R-loop signal around TTS (Figure 3D). Because R-loop formation is typically associated with specific genic characteristics such as gene expression level, torsional stress, and GC content (Chedin & Benham, 2020), we wondered if such genomic features were associated with an age-associated accumulation of R-loops over actively expressed genes. To test this, we clustered the genes from each heatmap from Figure 3D into four quartiles (Q1-4, with Q1 having the highest R-loop gain and Q4 having the highest R-loops loss), and assessed GC content, gene length and expression levels, respectively. Genes with TSS-enriched R-loops showed higher GC content than genes with R-loop losses (Wilcoxon-test, $p<2.2 \times 10^{-5}$), with no statistically significant association with long or highly expressed genes (Figure 3E-top). However, genes with TTS-enriched R-loops had high GC content, and were highly expressed and long (Kruskal-Wallis, $p<4.5 \times 10^{-2}$, $p<4.5 \times 10^{-4}$, and $p<2.1 \times 10^{-9}$, respectively) (Figure 3E-bottomC). Thus, these data show that age-associated accumulation of R-loops correlates with high GC content, length and expression levels.

3.2.4 Accumulation of R-loops in long genes correlates with decreased transcript levels in aging PRs

We previously showed that genes which decrease expression in aging PRs tend to be highly expressed and longer than genes that either increase or do not change expression (170). Moreover, aging *Drosophila* exhibit decrease in positive light response which correlates with decreased expression of long genes with neuronal function. Interestingly, recent reports have shown similar correlations between gene length and function in a variety of aging tissues and organisms, including humans (175,176). Because our current data showed that long genes accumulated R-loops with age, and R-loops can lead to RNA polymerase II arrest and transcription inhibition (177,178), we next investigated if there was an association between accumulation of R-loops and decreased gene expression during aging. Transcriptome profiling of PRs isolated from flies at day 10 and 50 revealed that 1700 genes (18%) were differentially expressed (DEG), with 722 genes (7%) decreasing expression and 978 genes (8%) increasing expression (p-adj<0.05, |FC|>1.5) (Figure 4A). To further evaluate the relationship between genic R-loop accumulation and gene expression, we identified R-loop containing genes (RCGs), by annotating high confidence peaks to the nearest TSS, and compared them to DEGs during aging. Venn diagram analysis revealed that 1388, or 69% genes were age-regulated at a transcript level and also contained at least one R-

loop (RCGs/DEGs) (Figure 4B). Further, gene length analysis of each category showed that RCG/DEGs were significantly longer than either DEGs or RCGs alone (Figure 4C), suggesting that accumulation of R-loops in long genes could contribute to gene expression changes during aging (Wilcoxon test, adjusted p-value < 2.22e-16).

To gain further insight into the biological processes associated with RCG/DEGs, we first used hierarchical clustering and identified 613 genes (46%) to be up-regulated and 775 genes (54%) being down-regulated. Importantly, we found that overall, nearly 80% of the genes that decreased expression during aging accumulated R-loops, while only 58% of the genes that increased expression during aging contained R-loops (Figure 4D). Gene Ontology (GO) term analysis revealed that down-regulated RCG/DEGs were enriched for functional and neuronal categories, including cognition and regulation of nervous system development, whereas up-regulated RCG/DEGs were enriched for metabolic processes, such as peptide biosynthesis and translation (Figure 4E). To further characterize the RCG/DEGs, we used literature mining tool BioLitMine and identified medical subject heading (MeSH) terms associated with RCG/DEGs that were significantly enriched for several eye- and brain-relevant diseases, such as epilepsy, eye abnormalities, and retinal and nerve degeneration (Figure 4F). Collectively, our data showed that R-loops accumulated at both age up- and down-regulated genes, suggesting that R-loops may contribute to gene expression regulation via multiple mechanisms. For example, transcription can be blocked by direct collision of RNAP by pre-formed R-loops from previous transcription rounds or alternatively, can be inhibited by an intrinsic R-loop formation in the wake of ongoing RNAP. However, increased R-loop levels in age up-regulated genes may be simply a result of increased expression of these genes. Importantly, given that R-loops accumulated at most of the age down-regulated genes, which are enriched for long genes with neuronal function, it suggests that R-loops may contribute to regulation of biological pathways relevant for eye-specific functions.

3.2.5 Top3β depletion in *Drosophila* PR neurons leads to increased R-loop levels and decreased visual function

The neuronal transcriptome is enriched for long and highly expressed genes, that undergo high level of torsional stress during transcription (179,180). To solve DNA and RNA topological problems, cells use conserved topoisomerase enzymes that play a critical role in a wide range of

fundamental metabolic processes in the genome (181,182). One of the enzymes is Top3β, a highly conserved, dual-activity topoisomerase in animals that can change the topology of both DNA and RNA (183) and unwind negatively supercoiled DNA that forms during transcription, an activity that prevents formation of R-loops (174). Loss of Top3β is associated with increased R-loop levels in mammalian cells (184) and has been shown to reduce lifespan in mice (167). In addition, mutations in Top3β are linked to neurological disorders, thus highlighting the critical role of Top3β in neuronal function (185). Interestingly, our recent proteomic study of the aging *Drosophila* eye (186) revealed a 20% decrease in Top3β protein levels (Figure 5A), suggesting that the aging eye might be sensitive to the loss of Top3β activity. Thus, we hypothesized that decreased Top3β levels could contribute to changes in R-loop homeostasis and neuronal function in aging PRs. To test this, we first depleted *Top3β* with ubiquitous RNAi in larvae (*tubP-Gal4>UAS-RNAi*) and measured bulk R-loop levels. Using DNA slot blot and RNA-DNA-specific antibody S9.6, we detected a 10% increase in R-loop levels in Top3β-depleted samples as compared to a control expressing non-specific RNAi (Figure 5B-C). Pre-treatment of DNA samples with RNase H1 led to a complete loss of the signal (Figure 5B-right), thus showing the specificity of the signal for RNA-DNA hybrids. In addition, qPCR analysis of Top3β transcript levels showed approximately 80% reduction in Top3β-depleted samples as compared to a control, thus validating the efficiency of the knockdown. Taken together, these data show that Top3β in *Drosophila* has a conserved role in maintenance of R-loop homeostasis.

Since loss of Top3β in *Drosophila* and mice leads to several neuronal phenotypes, such as disruption of synapse formation and behavioral impairments, we were next interested to see if depletion of Top3β specifically in PR neurons had any impact on visual function. Like most flying insects, *Drosophila* move towards light, thus exhibiting positive phototaxis (187). Importantly, we and others showed that positive phototaxis declines with age in flies (Carbone et al., 2016; Hall et al., 2017). To assess changes in visual behavior upon loss of Top3β, we depleted Top3β transcripts specifically in photoreceptors using *Rh1-Gal4>UAS-RNAi* and performed phototaxis assays at days 10 and 30 post-eclosion. As expected, there was approximately 15% decrease in positive phototaxis in the control flies between day 10 and day 30 (Figure 5D). Notably, while flies with PR-specific depletion of Top3β showed no significant difference in the phototactic response at day 10, they showed approximately 60% decrease in visual behavior at day 30 as compared to a control (Wilcoxon test, p-value<0.47, and <0.015, respectively). Importantly, this decrease in visual

behavior was not due to the loss of PR neurons, as optical neutralization showed no retinal degeneration in *Rh1-Gal4>UAS-RNAi* flies at day 30 post-eclosion (Figure 5E). Thus, our data show that Top3β is required for maintenance of proper visual function in aging *Drosophila* photoreceptor neurons.

3.2.6 Top3β regulates expression of a subset of long genes associated with neuronal function in photoreceptors

Given the role of Top3β in the resolution of torsional stress during transcription, we next hypothesized that Top3β might be required to regulate the expression of genes with neuronal function, which tend to be long and highly expresses (191). To test this hypothesis, we analyzed the transcriptome of PR neurons depleted for Top3β in *Rh1-Gal4>UAS-RNAi; UAS-GFP^{KASH}* flies at day 30 using our NIE protocol. Differential expression analysis using DESeq2 between Top3β-RNAi and control, revealed that expression of approximately 1% of genes was regulated by Top3β (66 out 6500, FDR<0.05) (Figure 6A). Additionally, quantitative analysis of gene length based on whether a gene was differentially expressed in Top3β depleted PRs revealed, that genes with decreased expression were highly and significantly enriched for long genes relative to genes that either increased or did not change expression (Figure 6B). These data thus suggest that transcriptional regulation of long genes in the eye is particularly sensitive to decreased Top3β levels. Gene Ontology (GO) enrichment analysis of all Top3β-dependent genes revealed that genes with decreased expression were highly enriched for genes with neuronal functions as shown by the gene concept network analysis (Cnetplot) (Figure 6C). These genes included *Tenascin major* (*Ten-m; FBgn0004449*) and *Tenascin accessory* (*Ten-a; FBgn0267001*), which form a transmembrane heterodimer involved in synapse regulation, *Tripartite motif containing 9* (*Trim9; FBgn0051721*), a E3 ubiquitin ligase involved in neurogenesis, axon guidance, and eye development, and *knockout* (*ko; FBgn0020294*), a storkhead-box protein involved in axon guidance. Importantly, GO term analysis of up-regulated genes did not lead to any significant biological category enrichment. Thus, our data show that in PR neurons, Top3β is required to maintain gene expression levels of long genes that are involved in neuronal function.

Next, we asked whether the expression of Top3β-dependent genes was mis-regulated during in aging PR neurons. To do this, we compared genes that were down-regulated either during aging

or upon depletion of Top3β. Venn diagram revealed that 30% of Top3β-dependent genes were also down-regulated during aging ("shared genes"), with the overlap being statistically significant (Figure 6D). Further, individual inspection of these shared genes revealed that majority of them were long and showed similar changes in fold change expression during aging and Top3β depletion. Thus, our data show that during aging, several long genes important for neuronal function decrease expression in a Top3β-dependent manner. Further, age- and Top3β-shared genes were significantly longer than Top3β-dependent genes (Figure 6E), suggesting that expression of longer genes is particularly sensitive to loss of Top3β during aging. However, gene length analysis of genes differentially expressed during aging in PRs revealed that down-regulated genes were significantly longer than up-regulated genes (Figure 6F-left). In addition, gene length analysis of our previously published RNA-seq data from aging eyes of Rh1>GFP^{KASH} flies (192) revealed a similar trend (Figure 6F-right) indicating that expression of the long genes in the aging retina is particularly sensitive to dysregulation of molecular mechanisms that include Top3β.

Collectively, our data showed that Top3β is required to maintain expression of a specific subset of genes with neuronal function that tend to be very long and thus are most likely sensitive to loss of topoisomerase activity due to high levels of torsional stress. This suggests that during aging, proper levels of Top3β are required to maintain R-loop homeostasis and expression of genes important for visual function.

3.2.7 Overexpression of Top3β or RNaseH1 enhances visual function during aging

To further validate the role of topoisomerase activity and maintenance of R-loop homeostasis in visual function during aging, we over-expressed either Top3β or human, nuclear localized, RNASEH1 in *Drosophila* eyes under the control of longGMR-GAL4 driver, which induces high expression in PRs and assessed visual behavior using phototaxis in 10-, 20-, and 30-day old flies. Importantly, we validated that RNASEH1 is expressed using qPCR and showed that bulk R-loop levels are decreased upon expression of RNASEH1 relative to a no driver control. In addition, we over-expressed Top3β under control of longGMR in a Top3β null background (Top3β$^{-/-}$; longGMR>Top3β). Our data showed that at day 10, over-expression of Top3β in the eyes of Top3β null flies resulted in similar light response as that of control flies (Rh1>siControl, see Figure 5D) or longGMR>siControl flies (192) (Figure 6G). Furthermore, Top3β$^{-/-}$; longGMR>Top3β flies

showed an age-associated decrease in positive light response which is consistent with control Rh1>siControl flies (Figure 6G-grey). Importantly, flies with Top3β over-expression (Top3β$^{+/+}$; longGMR>Top3β) showed an enhanced positive light response at either day 10 or day 20 (Figure 6G-light blue), corroborating that proper levels of Top3β contribute to visual response in flies. Moreover, over-expression of RNase H1 in *Drosophila* eyes (Top3β$^{+/+}$; longGMR>RNaseH1) showed enhanced visual response in 10- and 20-day old flies as compared to control flies, similar to that of flies with Top3β over-expression (Figure 6G). These data suggest that Top3β function and R-loop resolution by RNase H1 contribute to proper visual function in *Drosophila*. Notably, qPCR analysis performed in flies with RNase H1 over-expression showed no significant difference in Top3β transcript level as compared to that in wild type flies suggesting that the enhanced visual response in these flies is a result of increased R-loop removal. However, all tested genotypes showed decreased positive phototaxis at 30 days post-eclosion. Optical microscopy imaging of the eyes did not show any changes in the eye morphology in 30-day old flies and over-expression of Top3β using longGMR driver resulted in approximately 50% decrease in R-loop levels in eyes from 10-day old flies. These data imply that additional aging mechanisms contribute to regulation of visual function in the eye, possibly including R-loops accumulated as a result of aberrant expression of additional R-loop metabolism-associated factors

3.3   Discussion

While R-loops were previously considered to be mere byproducts of transcription, it has been demonstrated that R-loops play a significant physiological role in cellular biology of multiple organisms, including humans. Notably, there is a growing body of evidence that links R-loop accumulation to transcriptional imbalance and genomic instability, two main hallmarks of aging (1). Furthermore, dysregulation of R-loop homeostasis has been linked to human pathologies, including neurodegeneration (193). Since age is the main risk factor for many neurodegenerative diseases, our current study focused on characterizing the changes in R-loop landscape induced during aging and evaluating the impact of R-loops on the gene expression, specifically in *Drosophila* photoreceptor neurons.

Characterization of global R-loop levels in aging PRs showed an increase in R-loops by middle age, with an additional significant increase late in aging. To our knowledge, this is the first

published report demonstrating changes in R-loop levels in a specific tissue during aging. To further evaluate the R-loop distribution genome-wide, we modified a recently published R-loop mapping strategy, called MapR, coupled with high-throughput sequencing that can be useful when material is limiting. In addition, given that this method involves incubation of isolated nuclei/cells with a recombinant mutant form of RNase H1 tethered with MNase and therefore does not require modification of the organismal genome, it is well-suited for studies in whole animals. Using this approach, our data demonstrated that R-loops covered approximately 10% of the *Drosophila* PR genome, which is similar to the reported genomic distribution of R-loops in other organisms (194,195). Consistent with previous reports, R-loops were associated with known genomic hotspots such as gene termini and specific genic features such as high GC content, gene length and expression level. The aging transcriptomes of multiple organisms and cell types show a positive correlation between transcriptional downregulation and specific genic features, such as gene length and GC content (176) and R-loops are known to play a key physiological role in transcription regulation due to their presence at promoters and terminators, where they regulate transcription initiation and termination, respectively (161). Thus, R-loop accumulation over these genomic regions may be a conserved mechanism that contributes to gene expression regulation in multiple cell types, including neurons.

Further, we observed a significant increase in R-loop signal over gene bodies and importantly, age-associated broadening of R-loop peak signal, suggesting that aging neurons accumulate R-loops at higher rate or R-loops are more persistent and potentially extend with age. Transcriptome of neurons is biased for longer transcript relative to non-neuronal cell types (191). Notably, long genes accumulate high topological stress during transcription and loss of topoisomerase activity has been shown to preferentially inhibit expression of (López-Otín et al. 2013) long genes (179). Since age-associated R-loop gains were particularly localized at long and highly expressed genes, we sought to further explore the impact of decreased topoisomerase activity on the maintenance of photoreceptor neuron homeostasis. DNA/RNA topoisomerase Top3β dysfunction is associated with increased R-loop levels in mammalian cells and mutations in Top3β are linked to neurological disorders, thus highlighting the critical role of Top3β in neuronal function (185). Importantly, Top3β protein levels decrease in aging *Drosophila* eyes (186). Here, we demonstrate that normal Top3β levels are important for maintenance of neuronal function, as shown by an age-associated decrease in visual behavior upon photoreceptor-specific downregulation of Top3β. In addition,

depletion of Top3β in PRs lead to decreased expression of a subset of long genes with neuronal function. Moreover, genes down-regulated during Top3β depletion contained higher R-loop levels than genes down-regulated during both aging and Top3β depletion, suggesting that regulation of R-loop homeostasis and expression of long genes in PR neurons are highly sensitive to loss of Top3β function. Moreover, Top3β or RNase H1 over-expression in the eyes enhanced positive light response in flies and mitigated age-dependent loss of visual function in middle-aged flies. However, systematic and integrative studies should be used to characterize how processes associated with late aging impact the pathways associated with R-loop homeostasis maintenance. Collectively, our data suggest that Top3β may function in regulation of gene expression and maintenance of R-loop homeostasis in a subset of long genes required for neuronal function in aging photoreceptors. Thus, imbalance in R-loop homeostasis during aging could make postmitotic neurons particularly susceptible to dysregulation of gene expression and loss of function leading to increased risk of age-related neurodegeneration.

Aging is accompanied by elevated incidence of ocular diseases such as age-related macular degeneration and glaucoma, which exhibit characteristics of neurodegenerative diseases including loss of function and irreversible neuronal cell loss. How does aging impact development and progression of age-associated chronic diseases is one of the key questions in the biology of aging. Our current studies demonstrate a novel finding that *Drosophila* photoreceptor neurons progressively accumulate R-loops during aging, mostly at long and highly expressed genes. Importantly, integration of our transcriptomic and R-loop mapping data shows that majority of genes that decrease expression in aging PRs accumulate R-loops, thus suggesting that R-loops could be involved in cell physiology of aging neurons via inhibition of gene expression. Moreover, persistent formation of R-loops often leads to increased DNA damage, which is associated both with aging and neurodegeneration. Given that mutations in number of proteins involved in R-loop biology are implicated in neurodegenerative disease, our studies suggest that both aging and neurodegeneration may be sensitive to dysfunction in similar pathways.
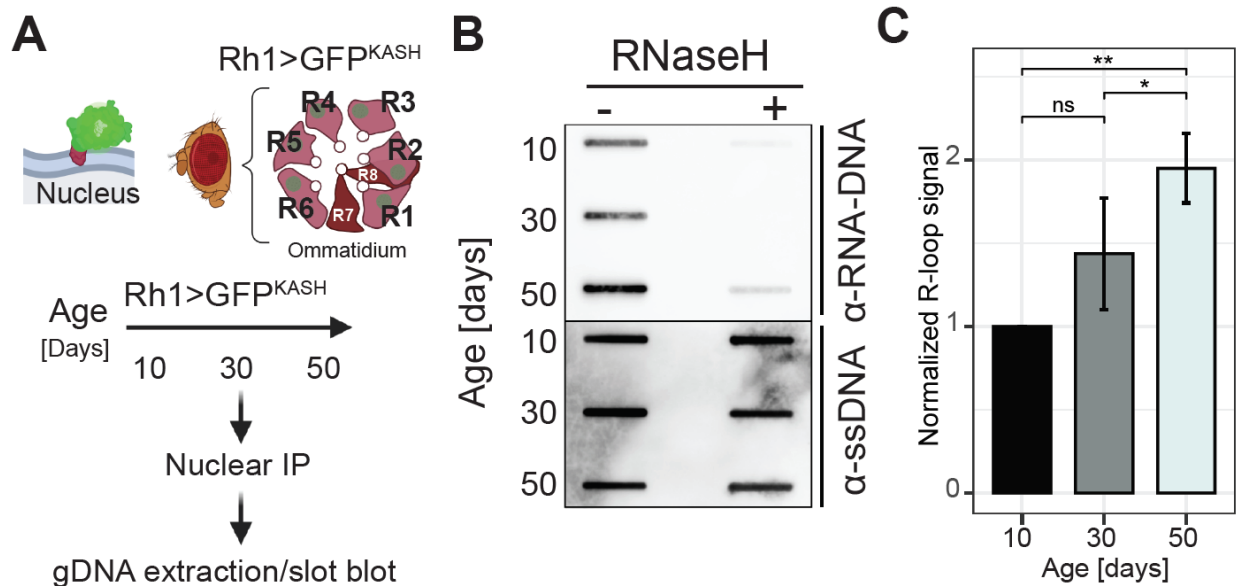
Figure 3.1 A Schematic of experimental outline to detect global levels of R-loops in aging photoreceptor neurons.

Top: diagram of the cellular localization of the GFP^{KASH} protein. Dark blue lines represent each lipid layer within the nuclear membrane. Bottom: diagram of an ommatidium, a structural subunit in the *Drosophila* compound eye. Each ommatidium is composed of 8 photoreceptor neurons, labeled R1 to R8. Outer photoreceptors (R1-R6) express the *ninaE* (Rh1) gene. B. Slot blot analysis of R-loop levels from photoreceptor nuclei at day 10, 30 and 50 post-eclosion treated with (right) or without (left) RNase H1. Slot blots were performed using RNA-DNA hybrid-specific S9.6 antibody (top) and ssDNA for loading control (bottom). C. Quantification of S9.6 slot blot signal in aging PRs from (b). Values above 1 represent increase signal relative to day 10. Mean +/- Standard Deviation (SD). p-value is obtained using t-test, (n=4).
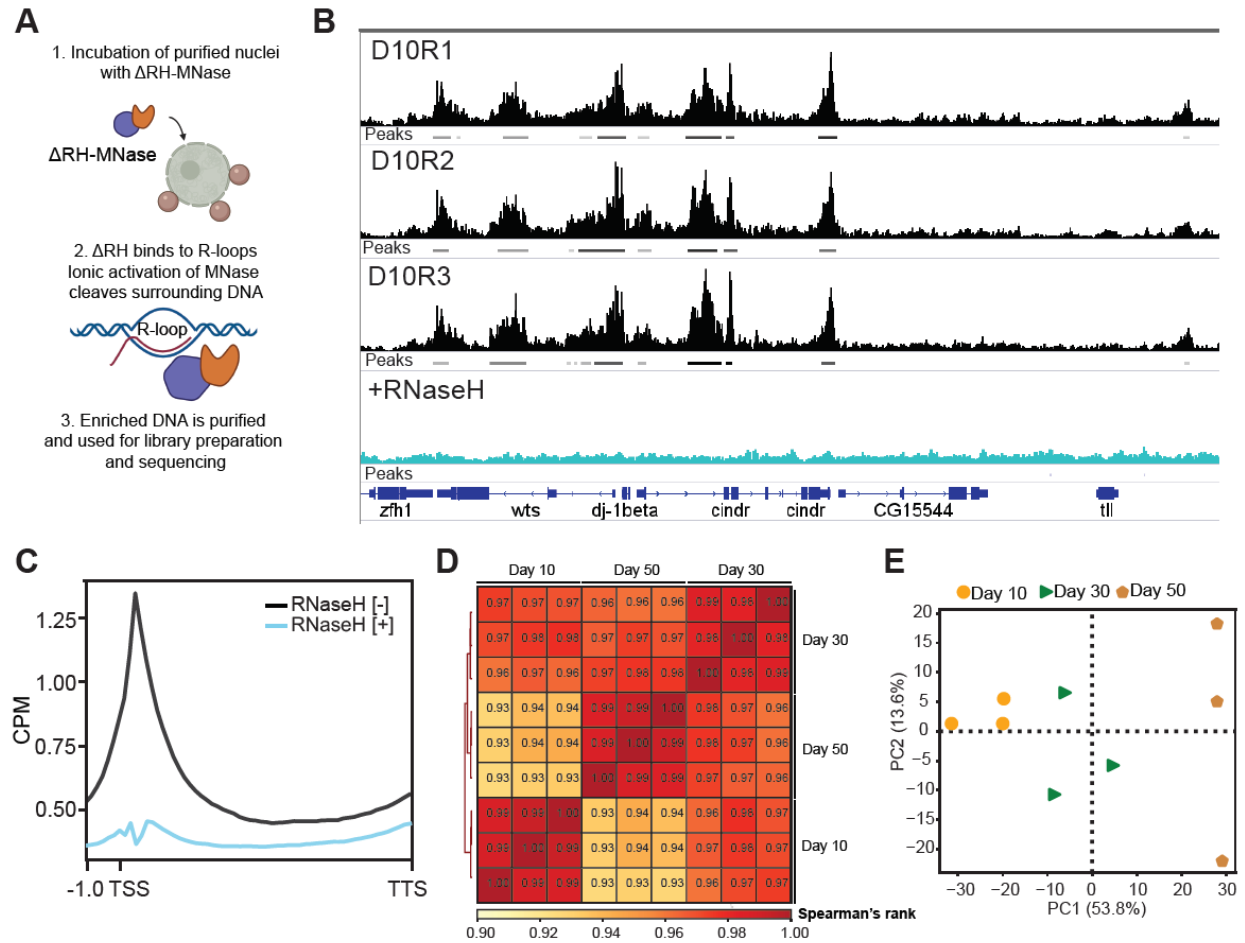
Figure 3.2 A. Schematic diagram of the R-loop mapping technique used in this study (MapR). Immuno-enriched nuclei are incubated with ΔRNaseH1-MNase (ΔRH-MNase), where ΔRH binds to R-loops. Ionic activation of MNase results in cleavage of surrounding DNA and subsequent R-loop enriched DNA release, which is purified and used for sequencing library preparation. B. Genome browser inspection of MapR track data on integrated genomic viewer (IgV) for a selected genomic region. Three independent biological replicates (R1-R3) from 10-day old flies' samples not pre-treated with RNaseH1 are shown in black, and a sample from nuclei that were pre-treated with RNaseH1 prior to MapR (see Methods) is shown in blue. Peaks obtained using MACS2 for each sample are also shown as bars under each corresponding sample track. C. Metaplot of CPM-normalized MapR signal over gene bodies for samples that were pre-treated with (blue) or without (black) RNase H1 prior to MapR (from b). D. Spearman correlation heatmap of Aging MapR read distribution over 1000-bp binned genome. Scores between 0 and 1 shown in each box correspond to Spearman's rank score. E. Principal component analysis (PCA) of Aging MapR samples based on read distribution over 1000-bp binned genome.
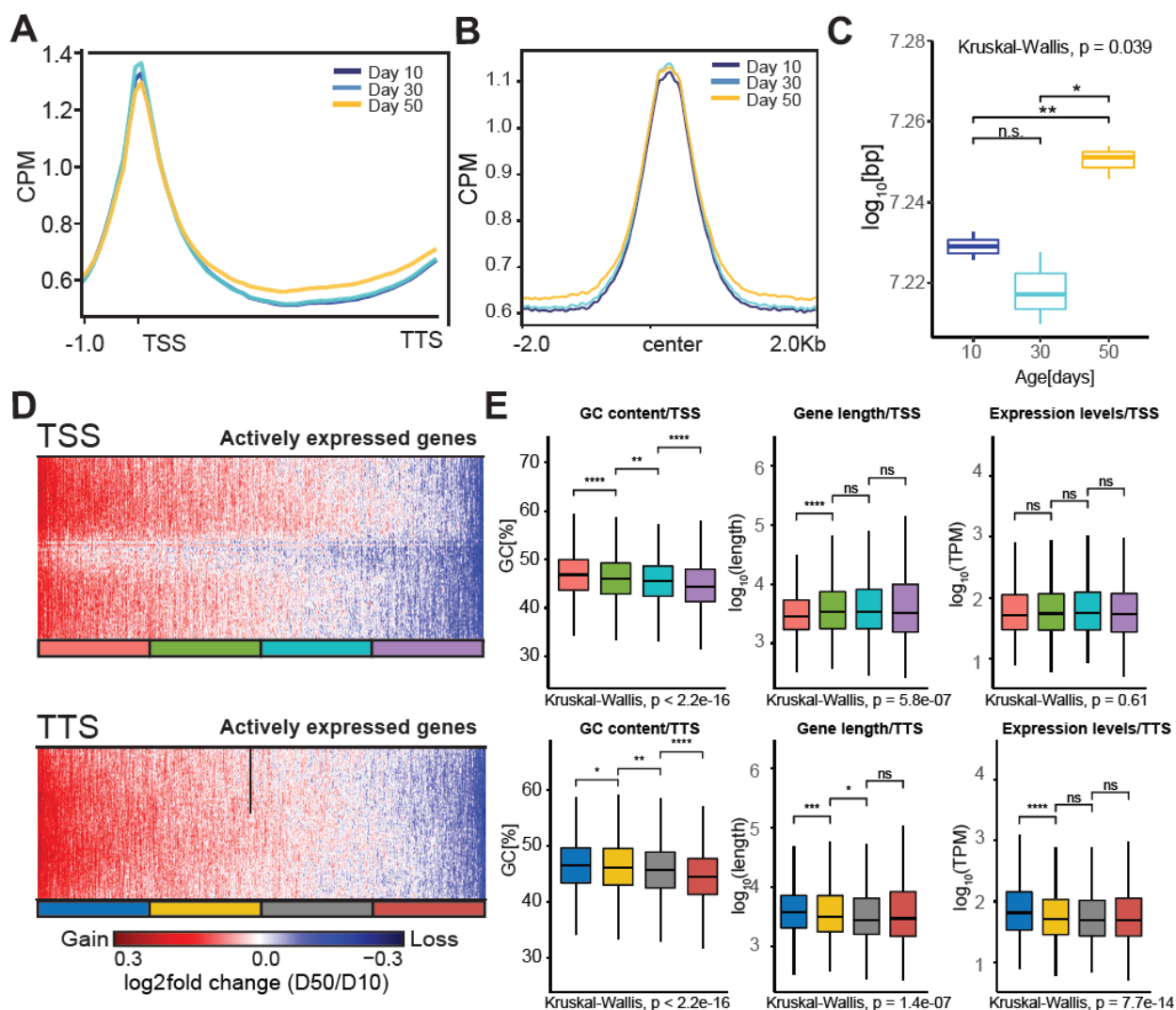
Figure 3.3 A. Metaplot of CPM-normalized MapR signal over gene bodies for all genes across age-timepoints. Signal is an average obtained from three independent biological replicates per age-timepoint. TSS indicates Transcription Start Site and TTS indicates Transcription Termination Site. B. Metaplot of CPM-normalized Aging MapR signal around peaks obtained using MACS2 during aging. C. Boxplot of genomic coverage of Aging MapR signal as defined by the total sum of peak width obtained at each time point. Peaks that mapped to scaffold or non-defined chromosomes were excluded from analysis. We used Wilcoxon Rank-Sum test to compare pair-wise differences in the distribution of genomic coverage amongst ages, (n=3). D. Heatmap showing log2 ratios of Aging MapR signal around the TSS (top) or TTS (bottom), comparing day 50 to day 10. Genes are ranked based on their fold change value and divided in four groups (quartiles) based on their position on the heatmap. E. Boxplot analysis of GC content, gene length and expression levels for each group of genes divided in four groups based on the Aging MapR fold changes around TSS (top) and TTS (bottom). p-value is obtained using Wilcoxon test.
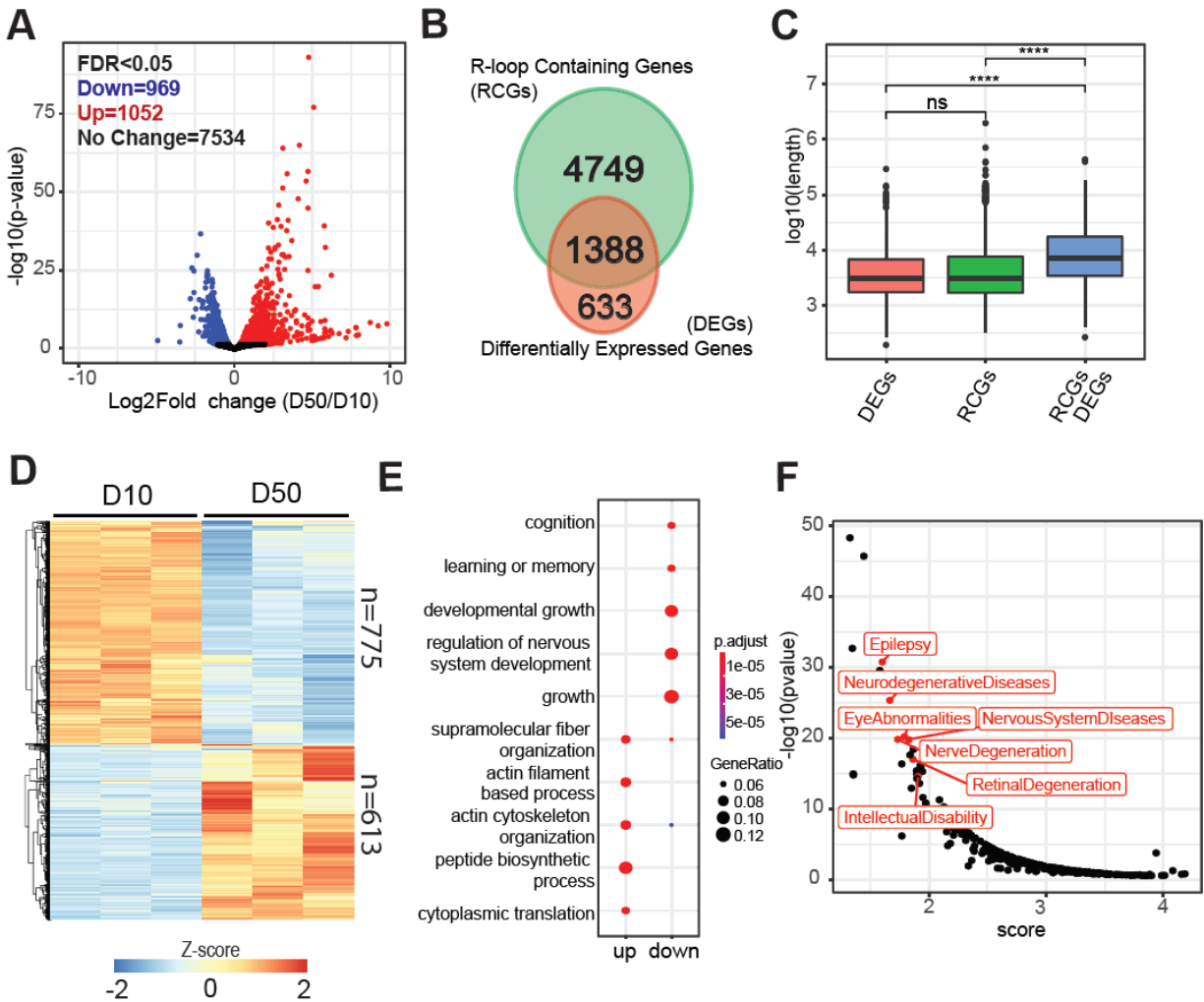
Figure 3.4 A. Volcano plot representing differentially expressed genes (DEGs) between day 50 and day 10. DEGs are obtained using DESeq2 (adjusted p-value < 0.05, |FC|>1.5). B. Venn diagram representing the overlap between R-loop containing genes (RCGs) and DEGs from (a). C. Box plot analysis of gene length for differentially expressed genes, R-loop containing genes, or RCG/DEGs from (b). p-value is obtained using Wilcoxon test. D. Hierarchically clustered heatmap of RNA-seq data for RCG/DEGs from (b). Normalized Z-scores are calculated based on normalized counts obtained using DESeq2, and the heatmap is divided into genes that were either up- or down-regulated with age. E. Dot plot of biological processes identified as significantly enriched in Gene Ontology (GO) term analysis for genes that were either up- or down-regulated from (d). F. Scatter plot depicting an enrichment analysis of diseases associated with genes that were down-regulated with age and contained at least one R-loop. Analysis performed using literature mining tool BioLitMine (197). A lower score (x-axis) represents higher enrichment.
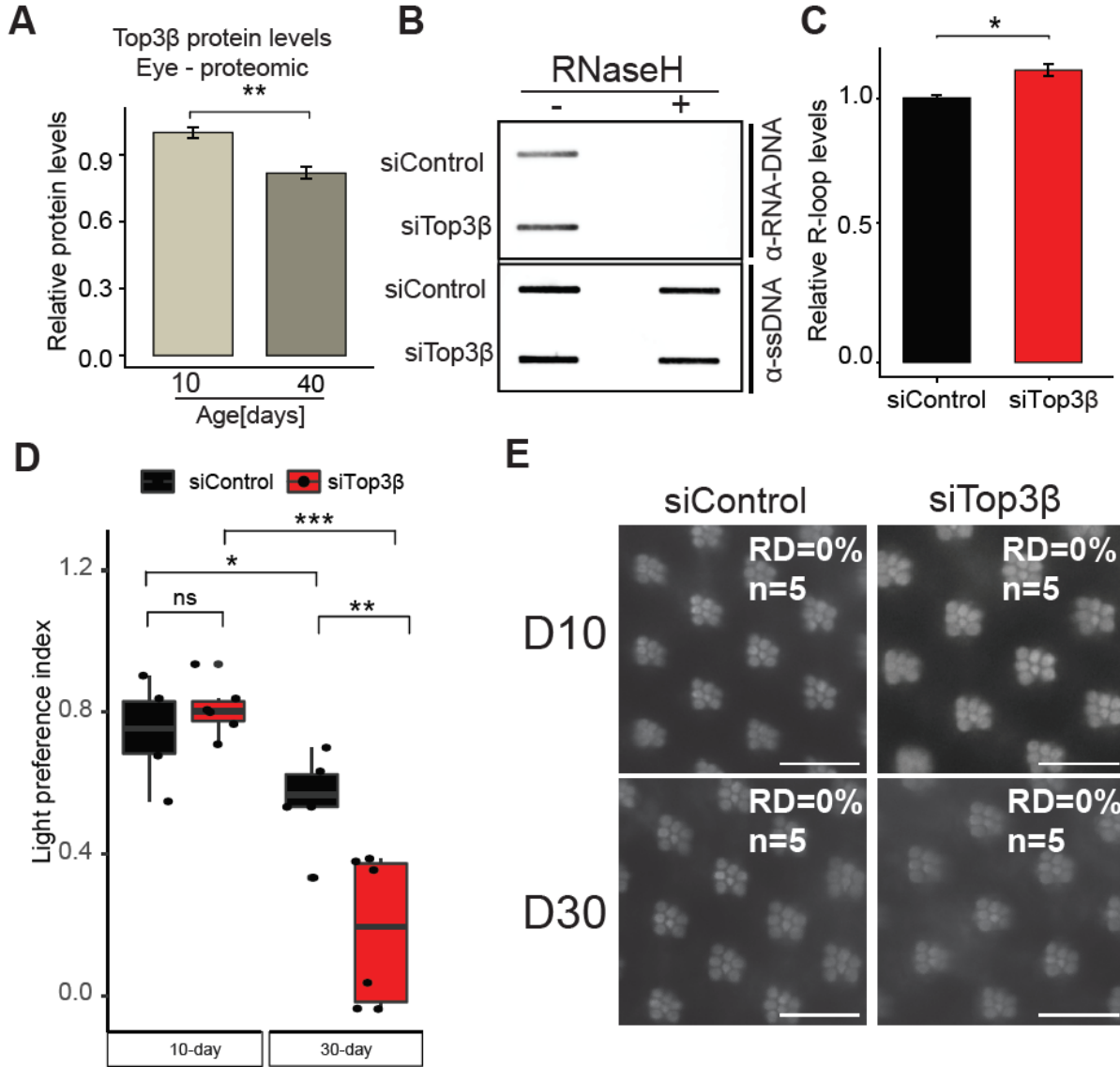
Figure 3.5 A Comparison of Top3β protein levels in aging eyes from Rh1>GFP^{KASH} flies, shown as normalized protein abundance. Proteomic samples were prepared from 10- and 40-day old flies, 100 eyes/sample (n=4). Raw data taken from (186). B. Slot blot analysis of R-loop levels from 3rd instar larvae ubiquitously expressing siRNA against mCherry (siControl) or against Top3β (siTop3β). Samples were treated with (right) or without (left) RNase H1. Slot blots were performed using S9.6 antibody (top) and ssDNA for loading control (bottom). C. Quantification of S9.6 slot blot from (b). S9.6 signal is normalized to ssDNA slot blot signal, (n=3). D. Box plots showing the light preference indices (positive phototaxis) for Rh1>GFP^{KASH}, mCherry-RNAi (siControl) or Rh1>GFP^{KASH}, Top3β-RNAi (siTop3β) at day 10 and 30 (6 biological replicates for each time point or RNAi, 27 - 33 male flies/replicate; total number of flies per fly strain=150-180). p value obtained using Wilcoxon test. E. Optic neutralization of siControl and siTop3β at day 10 and 30 post-eclosion from (d). Retinal degeneration (RD) scores were obtained by blindly quantifying 5 biological replicates. Score of 0% means there was no observable loss of rhabdomere or ommatidia.

Figure 3.6 A. Volcano plot representing differentially expressed genes (DEGs) between siTop3β and siControl-expressing photoreceptors at day 30 post eclosion. DEGs obtained using DESeq2 (adjusted p-value < 0.05, |FC|>1.5). Size of each point reflect the gene length of the whole gene as defined as most upstream TSS and most downstream TTS. B. Box plots showing the gene length (as $\log_2$-transformed bp) for genes identified as down-, up- or not regulated using DESeq2 in siTop3β photoreceptors relative to siControl (adjusted p-value < 0.05, |FC|>1.5). p-value obtained using Wilcoxon test. C. Gene concept network analysis (Cnetplot) of genes downregulated in siTop3β photoreceptors relative to siControl. Gene length in kilobases is shown next to each gene. D. Venn diagram representing the overlap of genes that were down-regulated in either aging (D50 vs D10) or upon loss of Top3β (siTop3β vs siControl). Overlap significance is denoted as a "overlapping p-value", obtained with a hypergeometric test. Odds ratio and Jaccard index are measurements of similarity. E. Box plots showing the gene length (as $\log_2$-transformed bp) for genes in the overlap identified in (d) or genes that were regulated by Top3β but not during aging. p-value is obtained using Wilcoxon test. F. Box plots showing the gene length (as $\log_{10}$-transformed bp) for genes that were identified as down- or up-regulated in either aging PRs (left) or eyes (right). Eye data was obtained from (192). G. Box plots showing the light preference indices (positive phototaxis) for Top3β-/-; longGMR>Top3β, Top3β-/-; longGMR>Top3β, and longGRM>RNaseH1 at day 10, 20, and 30 post-eclosion (6 biological replicates for sample; 29 - 31 male flies/experiment; total number of flies ~180). p-value obtained using Wilcoxon test.
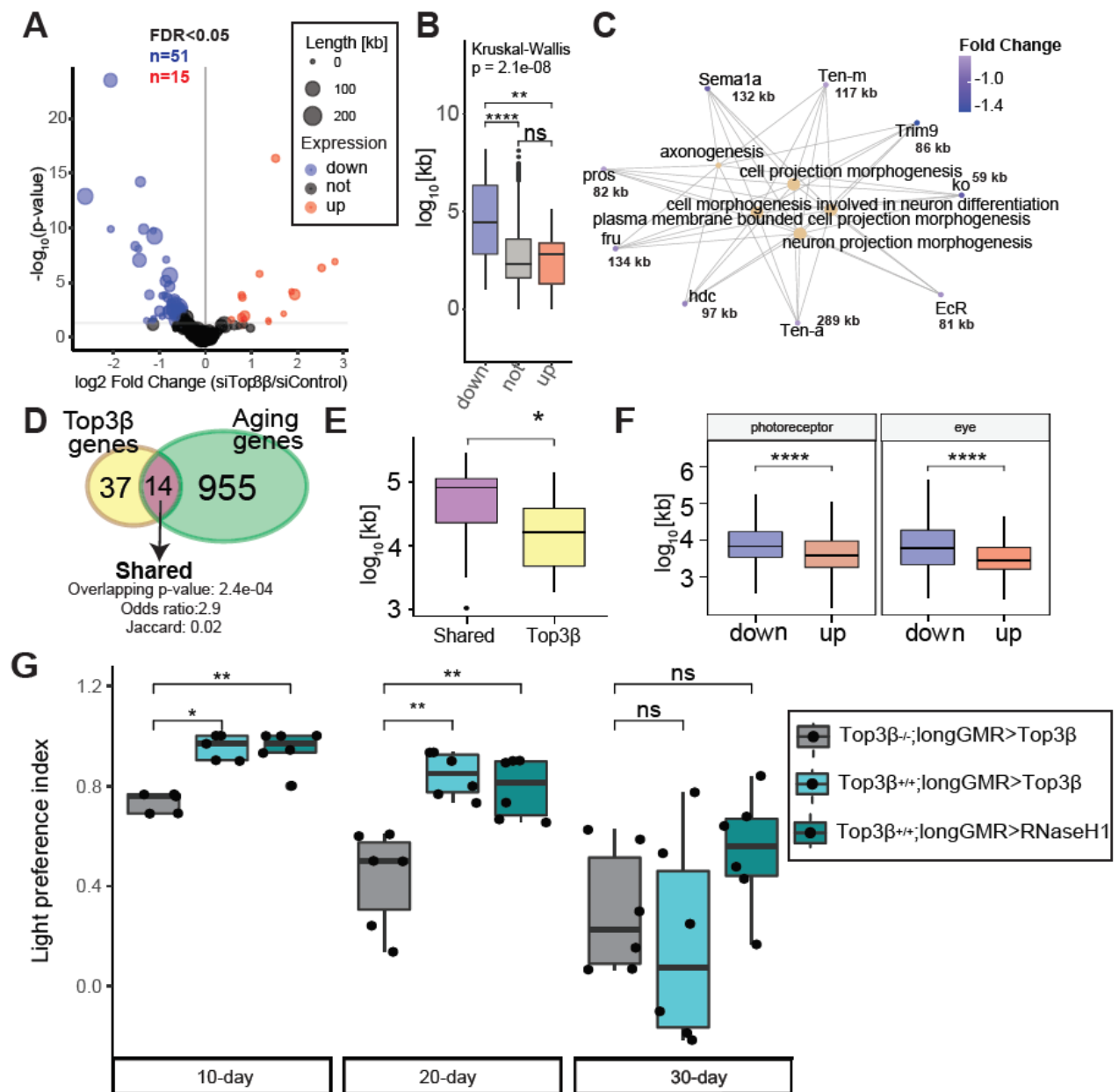
CHAPTER 4.　CONCLUSIONS

Here, I improved our previously published nuclear immuno-enrichment method to isolate cell-type specific nuclei. While I demonstrated that this protocol could be coupled with ChIP-seq and ATAC-seq to produce high-quality data, I found that CUT&Tag was not reproducible between technical replicates. Thus, in the future, it will be important to improve the CUT&Tag protocol. This will be important, since CUT&Tag has been shown to produce data of higher quality relative to traditional ChIP-seq. In addition, it will be important to establish if these nuclei immuno-enrichment protocol is amenable to single nuclei RNA-seq.

In addition, I discovered a role for Circadian-mediated transcription in maintaining visual health in *Drosophila*. In addition, I found that aging photoreceptors experience an age-associated increase in circadian transcription factor binding based on bioinformatic analysis. The caveat of our datasets is that samples were collected at one time point during the day. Since the transcription factor activity of circadian transcription factors changes during the day, one of the outstanding questions from my project is what are the specific changes in circadian, or rhythmic transcription that are observed during aging. To tackle this question, future studies in the lab will require to profile the nuclear transcriptome in a 24-hour cycle at different time points during age. Further, these studies should be couple with determining the genomic binding of circadian transcription factors (i.e. Clock and Cycle) via ChIP-seq or CUT&TAG. These sets of experiment will allow us to determine what are the age-associated changes in circadian transcription, and what biological processes in the retina are regulated in a circadian manner during aging. To further determine what are the molecular mechanisms that regulate CLK and CYC transcription factor activity, we will need to determine the genome-wide distribution of several factors involved in the core clock machinery that contribute to CLK and CYC binding, such as the CLK:CYC repressors TIM (*timeless),* PER (*period),* and CWO (*clockwork orange).* This is considered a promising avenue to follow up on since previous studies in other labs have shown decreased PER protein levels in aging photoreceptors. Under the rationale that less PER would lead to increased CLK:CYC binding to DNA, we hypothesized that decreased PER explains the increased CLK:CYC TF activity that we observed during aging.

Further, our studies also uncovered a role of circadian transcription in preventing retinal degeneration. Preliminary data in our lab shows that RNAi-mediated downregulation of transcript levels of histone methyltransferases involved in adding a third methyl group to Histone H3 Lysine 4 (H3K4) and Lysine 36 tri-methylation (H3K36), Set1 and Set2, respectively (Fig 4A), increased the risk of retinal degeneration with age (Escobedo Spencer, unpublished, not shown). Histone methylation is critical for maintenance of transcriptional outputs by regulating recruitment of splicing factors, or maintaining transcriptional fidelity. Further, other labs have shown that histone methylation is required to regulate rhythmic transcription. Interestingly, when I performed ChIP-seq of histone modifications H3K4me3 and H3K36me3 in aging photoreceptor neurons, I observed a widespread decrease in ChIP-seq signal over actively transcribed genes (Figure 4B). These data suggest that active histone modifications contribute to the aging transcriptome and are required to maintain circadian transcription. Given that loss of circadian rhythms and loss of histone modifications leads to retinal degeneration, it will be important to determine how loss of histone methylation (i.e. using RNAi) affects circadian transcription, and whether restoring histone methylation levels with age prevents retinal generation.
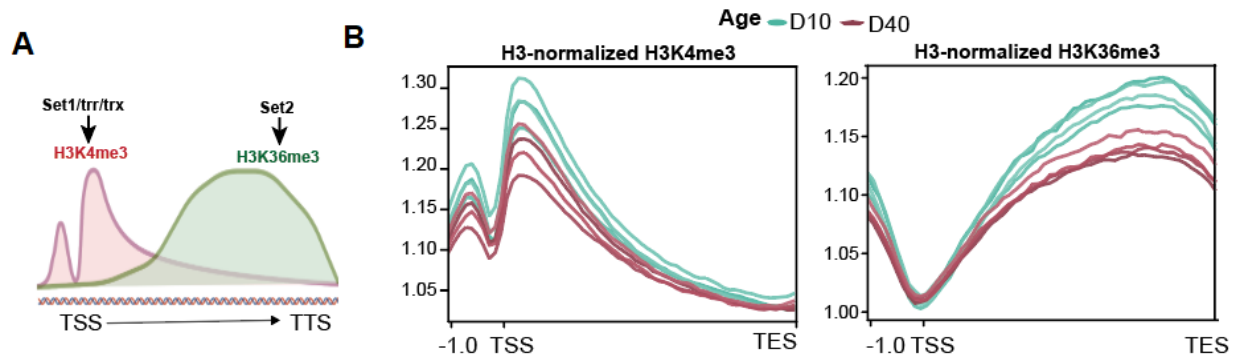
Figure 2. Aging photoreceptors neurons experience an age-associated decrease in histone methylation of Histone H3 lysine 4 (K4) and lysine 36 (K36).

A. In *Drosophila,* H3K4me3 can be deposited by three histone modifiers, Set1, trr, and trx. However, Set1 is the enzyme involved in deposition of bulk H3K4me3. H3K36me3 however, is deposited exclusively by Set2. B. Gene metaplots of H3-normalized H3K4me3 and H3K36me3 signal over actively expressed genes. Four independent biological replicates were assessed. Young (or 10-day old flies) are colored in green and old (or 40-day old flies) are colored in red.

REFERENCES

1. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. Cell. 2013 Jun 6;153(6):1194–217.

2. Stegeman R, Weake VM. Transcriptional Signatures of Aging. J Mol Biol. 2017 04;429(16):2427–37.

3. Bou Sleiman M, Jha P, Houtkooper R, Williams RW, Wang X, Auwerx J. The Gene-Regulatory Footprint of Aging Highlights Conserved Central Regulators. Cell Reports. 2020 Sep 29;32(13):108203.

4. Todhunter ME, Sayaman RW, Miyano M, LaBarge MA. Tissue aging: the integration of collective and variant responses of cells to entropic forces over time. Curr Opin Cell Biol. 2018 Oct;54:121–9.

5. Jauregui-Lozano J, Bakhle K, Weake VM. In vivo tissue-specific chromatin profiling in Drosophila melanogaster using GFP-tagged nuclei. Genetics [Internet]. 2021 Jul 1 [cited 2021 Nov 23];218(3). Available from: https://doi.org/10.1093/genetics/iyab079

6. Bonnel S, Mohand-Said S, Sahel JA. The aging of the retina. Experimental Gerontology. 2003 Aug 1;38(8):825–31.

7. Campello L, Singh N, Advani J, Mondal AK, Corso-Diaz X, Swaroop A. Aging of the Retina: Molecular and Metabolic Turbulences and Potential Interventions. Annu Rev Vis Sci. 2021 Jun 1;

8. Lu Y, Brommer B, Tian X, Krishnan A, Meer M, Wang C, et al. Reprogramming to recover youthful epigenetic information and restore vision. Nature. 2020 Dec;588(7836):124–9.

9. Yao K, Qiu S, Wang YV, Park SJH, Mohns EJ, Mehta B, et al. Restoration of vision after de novo genesis of rod photoreceptors in mammalian retinas. Nature. 2018 Aug;560(7719):484–8.

10. Stegeman R, Hall H, Escobedo SE, Chang HC, Weake VM. Proper splicing contributes to visual function in the aging Drosophila eye. Aging Cell [Internet]. 2018 Oct [cited 2020 Nov 2];17(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6156539/

11. Cowan CS, Renner M, De Gennaro M, Gross-Scherf B, Goldblum D, Hou Y, et al. Cell Types of the Human Retina and Its Organoids at Single-Cell Resolution. Cell. 2020 Sep 17;182(6):1623-1640.e34.

12. Ma J, Weake VM. Affinity-based Isolation of Tagged Nuclei from Drosophila Tissues for Gene Expression Analysis. JoVE (Journal of Visualized Experiments). 2014 Mar 25;(85):e51418.

13. Hall H, Medina P, Cooper DA, Escobedo SE, Rounds J, Brennan KJ, et al. Transcriptome profiling of aging Drosophila photoreceptors reveals gene expression trends that correlate with visual senescence. BMC Genomics. 2017 Nov 21;18(1):894.

14. Hall H, Ma J, Shekhar S, Leon-Salas WD, Weake VM. Blue light induces a neuroprotective gene expression program in Drosophila photoreceptors. BMC Neuroscience. 2018 Jul 20;19(1):43.

15. Slankster E, Kollala S, Baria D, Dailey-Krempel B, Jain R, Odell SR, et al. Mechanism underlying starvation-dependent modulation of olfactory behavior in Drosophila larva. Scientific Reports. 2020 Feb 20;10(1):3119.

16. Brown B, Mitra S, Roach FD, Vasudevan D, Ryoo HD. The Transcription Factor Xrp1 is Required for PERK- Mediated Antioxidant Gene Induction in Drosophila [Internet]. 2021 Sep [cited 2021 Oct 12] p. 2021.09.20.461097. Available from: https://www.biorxiv.org/content/10.1101/2021.09.20.461097v1

17. Huang HW, Ryoo HD. Drosophila fabp is a retinoid-inducible gene required for Rhodopsin-1 homeostasis and photoreceptor survival [Internet]. 2021 Apr [cited 2021 Oct 12] p. 2021.04.16.440122. Available from: https://www.biorxiv.org/content/10.1101/2021.04.16.440122v1

18. Loker R, Sanner JE, Mann RS. Cell-type-specific Hox regulatory strategies orchestrate tissue identity. Curr Biol. 2021 Jul 31;S0960-9822(21)00985-4.

19. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics. 2019 Apr;20(4):207–20.

20. Brahma S, Henikoff S. Epigenome Regulation by Dynamic Nucleosome Unwrapping. Trends in Biochemical Sciences. 2020 Jan 1;45(1):13–26.

21. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol. 2018 Apr;19(4):229–44.

22. Stadler J, Richly H. Regulation of DNA Repair Mechanisms: How the Chromatin Environment Regulates the DNA Damage Response. Int J Mol Sci [Internet]. 2017 Aug 5 [cited 2021 Mar 9];18(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5578105/

23. Hales KG, Korey CA, Larracuente AM, Roberts DM. Genetics on the Fly: A Primer on the Drosophila Model System. Genetics. 2015 Nov;201(3):815–42.

24. Bolus H, Crocker K, Boekhoff-Falk G, Chtarbanova S. Modeling Neurodegenerative Disorders in Drosophila melanogaster. Int J Mol Sci [Internet]. 2020 Apr 26 [cited 2021 Feb 26];21(9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246467/

25. Graham P, Pick L. Drosophila as a Model for Diabetes and Diseases of Insulin Resistance. Curr Top Dev Biol. 2017;121:397–419.

26. Piper MDW, Partridge L. Drosophila as a model for ageing. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2018 Sep 1;1864(9, Part A):2707–17.

27. Ugur B, Chen K, Bellen HJ. Drosophila tools and assays for the study of human diseases. Dis Model Mech. 2016 Mar 1;9(3):235–44.

28. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell. 2018 Apr 5;173(2):371-385.e18.

29. Lardenoije R, Iatrou A, Kenis G, Kompotis K, Steinbusch HWM, Mastroeni D, et al. The epigenetics of aging and neurodegeneration. Prog Neurobiol. 2015 Aug;131:21–64.

30. Chitikova Z, Steiner FA. Cell type-specific epigenome profiling using affinity-purified nuclei. genesis. 2016;54(4):160–9.

31. Bonn S, Zinzen RP, Perez-Gonzalez A, Riddell A, Gavin AC, Furlong EEM. Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. Nat Protoc. 2012 Apr 26;7(5):978–94.

32. Deal RB, Henikoff S. A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell. 2010 Jun 15;18(6):1030–40.

33. Maher KA, Bajic M, Kajala K, Reynoso M, Pauluzzi G, West DA, et al. Profiling of Accessible Chromatin Regions across Multiple Plant Species and Cell Types Reveals Common Gene Regulatory Principles and New Control Modules. Plant Cell. 2018 Jan;30(1):15–36.

34. Sijacic P, Bajic M, McKinney EC, Meagher RB, Deal RB. Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. Plant J. 2018 Apr;94(2):215–31.

35. Agrawal P, Chung P, Heberlein U, Kent C. Enabling cell-type-specific behavioral epigenetics in Drosophila: a modified high-yield INTACT method reveals the impact of social environment on the epigenetic landscape in dopaminergic neurons. BMC Biology. 2019 Apr 10;17(1):30.

36. Bozek M, Cortini R, Storti AE, Unnerstall U, Gaul U, Gompel N. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the Drosophila blastoderm. Genome Res [Internet]. 2019 Apr 8 [cited 2019 Dec 17]; Available from: http://genome.cshlp.org/content/early/2019/04/19/gr.242362.118

37. Henry GL, Davis FP, Picard S, Eddy SR. Cell type–specific genomics of Drosophila neurons. Nucleic Acids Res. 2012 Oct;40(19):9691–704.

38.   Jones SG, Nixon KCJ, Chubak MC, Kramer JM. Mushroom Body Specific Transcriptome Analysis Reveals Dynamic Regulation of Learning and Memory Genes After Acquisition of Long-Term Courtship Memory in Drosophila. G3 (Bethesda). 2018 Aug 29;8(11):3433–46.

39.   Amin NM, Greco TM, Kuchenbrod LM, Rigney MM, Chung MI, Wallingford JB, et al. Proteomic profiling of cardiac tissue by isolation of nuclei tagged in specific cell types (INTACT). Development. 2014 Feb 15;141(4):962–73.

40.   Ambati S, Yu P, McKinney EC, Kandasamy MK, Hartzell D, Baile CA, et al. Adipocyte nuclei captured from VAT and SAT. BMC Obes [Internet]. 2016 Jul 19 [cited 2021 Mar 6];3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4949929/

41.   Lam KWG, Brick K, Cheng G, Pratto F, Camerini-Otero RD. Cell-type-specific genomics reveals histone modification dynamics in mammalian meiosis. Nat Commun [Internet]. 2019 Aug 23 [cited 2021 Apr 27];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6707301/

42.   Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993 Jun 1;118(2):401–15.

43.   Yu J, Starr DA, Wu X, Parkhurst SM, Zhuang Y, Xu T, et al. The KASH domain protein MSP-300 plays an essential role in nuclear anchoring during Drosophila oogenesis. Developmental Biology. 2006 Jan 15;289(2):336–45.

44.   Ma J, Brennan KJ, D'Aloia MR, Pascuzzi PE, Weake VM. Transcriptome Profiling Identifies Multiplexin as a Target of SAGA Deubiquitinase Activity in Glia Required for Precise Axon Guidance During Drosophila Visual Development. G3 (Bethesda). 2016 Jun 1;6(8):2435–45.

45.   Mollereau B, Wernet MF, Beaufils P, Killian D, Pichaud F, Kühnlein R, et al. A green fluorescent protein enhancer trap screen in Drosophila photoreceptor cells. Mechanisms of Development. 2000 May 1;93(1):151–60.

46.   Fischer JA, Acosta S, Kenny A, Cater C, Robinson C, Hook J. Drosophila Klarsicht Has Distinct Subcellular Localization Domains for Nuclear Envelope and Microtubule Localization in the Eye. Genetics. 2004 Nov 1;168(3):1385–93.

47.   Shaiken TE, Opekun AR. Dissecting the cell to nucleus, perinucleus and cytosol. Scientific Reports. 2014 May 12;4(1):4923.

48.   Rasch EM, Barr HJ, Rasch RW. The DNA content of sperm of Drosophila melanogaster. Chromosoma. 1971 Mar 1;33(1):1–18.

49.   Göpfert MC, Robert D. Turning the key on Drosophila audition. Nature. 2001 Jun;411(6840):908–908.

50.   Stark WS, Thomas CF. Microscopy of multiple visual receptor types in Drosophila. Mol Vis. 2004 Dec 15;10:943–55.

51.   Helfrich-Förster C, Homberg U. Pigment-dispersing hormone-immunoreactive neurons in the nervous system of wild-type Drosophila melanogaster and of several mutants with altered circadian rhythmicity. Journal of Comparative Neurology. 1993;337(2):177–90.

52.   Nichols R, Schneuwly SA, Dixon JE. Identification and characterization of a Drosophila homologue to the vertebrate neuropeptide cholecystokinin. J Biol Chem. 1988 Sep 5;263(25):12167–70.

53.   Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nature Methods. 2017 Oct;14(10):959–62.

54.   Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012 Sep;22(9):1813–31.

55.   Bai L, Morozov AV. Gene regulation by nucleosome positioning. Trends Genet. 2010 Nov;26(11):476–83.

56.   Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. 2008 Jan 23;27(2):406–20.

57.   Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. Mol Cell Biol. 2016 Feb 16;36(5):662–7.

58.   Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, et al. Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. Cell Reports. 2014 Nov 6;9(3):1163–70.

59.   Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun [Internet]. 2019 Apr 29 [cited 2021 Feb 5];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6488672/

60.   Davie K, Jacobs J, Atkins M, Potier D, Christiaens V, Halder G, et al. Discovery of Transcription Factors and Regulatory Regions Driving In Vivo Tumor Development by ATAC-seq and FAIRE-seq Open Chromatin Profiling. PLoS Genet [Internet]. 2015 Feb 13 [cited 2021 Mar 3];11(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4334524/

61.   Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213–8.

62.   Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Reinberg D, editor. eLife. 2017 Jan 12;6:e21856.

63.   Potter CJ, Tasic B, Russler EV, Liang L, Luo L. The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. Cell. 2010 Apr 30;141(3):536–48.

64.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

65.   Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan;29(1):15–21.

66.   Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012 Apr;9(4):357–9.

67.   Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

68.   Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014 Jul 1;42(Web Server issue):W187–91.

69.   Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS. 2012 May;16(5):284–7.

70.   Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res. 2018 Sep 17;7:1338.

71.   Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

72.   Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013 May;41(10):e108.

73.   Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015 Jul 15;31(14):2382–3.

74.   López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. Cell. 2013 Jun 6;153(6):1194–217.

75.   Parapuram SK, Cojocaru RI, Chang JR, Khanna R, Brooks M, Othman M, et al. Distinct signature of altered homeostasis in aging rod photoreceptors: implications for retinal diseases. PLoS One. 2010 Nov 8;5(11):e13885.

76.   Gemenetzi M, Lotery AJ. Epigenetics in age-related macular degeneration: new discoveries and future perspectives. Cell Mol Life Sci. 2020 Mar;77(5):807–18.

77. Li X, He S, Zhao M. An Updated Review of the Epigenetic Mechanism Underlying the Pathogenesis of Age-related Macular Degeneration. Aging Dis. 2020 Oct;11(5):1219–34.

78. Martins R, Lithgow GJ, Link W. Long live FOXO: unraveling the role of FOXO proteins in aging and longevity. Aging Cell. 2016 Apr;15(2):196–207.

79. Murtaza G, Khan AK, Rashid R, Muneer S, Hasan SMF, Chen J. FOXO Transcriptional Factors and Long-Term Living. Oxid Med Cell Longev [Internet]. 2017 [cited 2021 Apr 26];2017. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5574317/

80. Sun X, Chen WD, Wang YD. DAF-16/FOXO Transcription Factor in Aging and Longevity. Front Pharmacol [Internet]. 2017 Aug 23 [cited 2021 Apr 26];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5572328/

81. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol. 2020 Feb 3;21:22.

82. Buono L, Martinez-Morales JR. Retina Development in Vertebrates: Systems Biology Approaches to Understanding Genetic Programs: On the Contribution of Next-Generation Sequencing Methods to the Characterization of the Regulatory Networks Controlling Vertebrate Eye Development. Bioessays. 2020 Apr;42(4):e1900187.

83. Potier D, Davie K, Hulselmans G, Naval Sanchez M, Haagen L, Huynh-Thu VA, et al. Mapping gene regulatory networks in Drosophila eye development by large-scale transcriptome perturbations and motif inference. Cell Rep. 2014 Dec 24;9(6):2290–303.

84. Wang J, Zibetti C, Shang P, Sripathi SR, Zhang P, Cano M, et al. ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. Nat Commun [Internet]. 2018 Apr 10 [cited 2021 Mar 3];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5893535/

85. Rister J, Desplan C. The retinal mosaics of opsin expression in invertebrates and vertebrates. Developmental Neurobiology. 2011;71(12):1212–26.

86. Montell C. Drosophila visual transduction. Trends in Neurosciences. 2012 Jun 1;35(6):356–63.

87. Jauregui-Lozano J, Bakhle K, Weake VM. In vivo tissue-specific chromatin profiling in Drosophila melanogaster using GFP-tagged nuclei. Genetics [Internet]. 2021 May 22 [cited 2021 May 22];(iyab079). Available from: https://doi.org/10.1093/genetics/iyab079

88. Grandison RC, Wong R, Bass TM, Partridge L, Piper MDW. Effect of a Standardised Dietary Restriction Protocol on Multiple Laboratory Strains of Drosophila melanogaster. PLOS ONE. 2009 Jan 1;4(1):e4067.

89. Francois M, Donovan P, Fontaine F. Modulating transcription factor activity: Interfering with protein-protein interaction networks. Seminars in Cell & Developmental Biology. 2020 Mar 1;99:12–9.

90.   Filtz TM, Vogel WK, Leid M. Regulation of transcription factor activity by interconnected, post-translational modifications. Trends Pharmacol Sci. 2014 Feb;35(2):76–85.

91.   Hall H, Cooper BR, Qi G, Wijeratne AB, Mosley AL, Weake VM. Quantitative Proteomic and Metabolomic Profiling Reveals Altered Mitochondrial Metabolism and Folate Biosynthesis Pathways in the Aging Drosophila Eye. Molecular & Cellular Proteomics. 2021 Jan 1;20:100127.

92.   Berest I, Arnold C, Reyes-Palomares A, Palla G, Rasmussen KD, Giles H, et al. Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. Cell Reports. 2019 Dec 3;29(10):3147-3159.e12.

93.   Hammelman J, Patel T, Closser M, Wichterle H, Gifford D. Ranking Reprogramming Factors for Directed Differentiation. bioRxiv. 2021 May 17;2021.05.14.444080.

94.   Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014 Sep 11;158(6):1431–43.

95.   Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010 May 28;38(4):576–89.

96.   Vinayagam A, Kulkarni MM, Sopko R, Sun X, Hu Y, Nand A, et al. An Integrative Analysis of the InR/PI3K/Akt Network Identifies the Dynamic Response to Insulin Signaling. Cell Reports. 2016 Sep 13;16(11):3062–74.

97.   Shokri L, Inukai S, Hafner A, Weinand K, Hens K, Vedenko A, et al. A Comprehensive Drosophila melanogaster Transcription Factor Interactome. Cell Reports. 2019 Apr 16;27(3):955-970.e7.

98.   Haigis MC, Yankner BA. The Aging Stress Response. Mol Cell. 2010 Oct 22;40(2):333–44.

99.   Zhang K, Chaillet JR, Perkins LA, Halazonetis TD, Perrimon N. Drosophila homolog of the mammalian jun oncogene is expressed during embryonic development and activates transcription in mammalian cells. Proc Natl Acad Sci U S A. 1990 Aug;87(16):6281–5.

100. Dekanty A, Lavista-Llanos S, Irisarri M, Oldham S, Wappner P. The insulin-PI3K/TOR pathway induces a HIF-dependent transcriptional response in Drosophila by promoting nuclear localization of HIF-α/Sima. Journal of Cell Science. 2005 Dec 1;118(23):5431–41.

101. Baird NA, Turnbull DW, Johnson EA. Induction of the Heat Shock Pathway during Hypoxia Requires Regulation of Heat Shock Factor by Hypoxia-inducible Factor-1*. Journal of Biological Chemistry. 2006 Dec 15;281(50):38675–81.

102. Chakrabarti S, Poidevin M, Lemaitre B. The Drosophila MAPK p38c Regulates Oxidative Stress and Lipid Homeostasis in the Intestine. PLOS Genetics. 2014 Sep 25;10(9):e1004659.

103. Sano Y, Akimaru H, Okamura T, Nagao T, Okada M, Ishii S. Drosophila Activating Transcription Factor-2 Is Involved in Stress Response via Activation by p38, but Not c-Jun NH2-Terminal Kinase. MBoC. 2005 Jun 1;16(6):2934–46.

104. Hall BS, Barnett YA, Crofts JJ, Chuzhanova N. Identification of novel genes associated with longevity in Drosophila melanogaster - a computational approach. Aging (Albany NY). 2019 Dec 3;11(23):11244–67.

105. Patke A, Young MW, Axelrod S. Molecular mechanisms and physiological importance of circadian rhythms. Nat Rev Mol Cell Biol. 2020 Feb;21(2):67–84.

106. Menet JS, Hardin PE. Circadian Clocks: The tissue is the issue. Curr Biol. 2014 Jan 6;24(1):R25–7.

107. Giebultowicz JM, Long DM. Ageing and Circadian rhythms. Curr Opin Insect Sci. 2015 Feb 1;7:82–6.

108. Hood S, Amir S. The aging clock: circadian rhythms and later life. J Clin Invest. 2017 Feb 1;127(2):437–46.

109. Trott AJ, Menet JS. Regulation of circadian clock transcriptional output by CLOCK:BMAL1. PLoS Genet. 2018 Jan;14(1):e1007156.

110. Hardin PE. Transcription Regulation within the Circadian Clock: The E-box and Beyond. J Biol Rhythms. 2004 Oct 1;19(5):348–60.

111. Tanoue S, Krishnan P, Krishnan B, Dryer SE, Hardin PE. Circadian Clocks in Antennal Neurons Are Necessary and Sufficient for Olfaction Rhythms in Drosophila. Current Biology. 2004 Apr 20;14(8):638–49.

112. Mahesh G, Rivas GBS, Caster C, Ost EB, Amunugama R, Jones R, et al. Proteomic analysis of Drosophila CLOCK complexes identifies rhythmic interactions with SAGA and Tip60 complex component NIPPED-A. Sci Rep [Internet]. 2020 Oct 21 [cited 2021 Mar 24];10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7578830/

113. Richier B, Michard-Vanhée C, Lamouroux A, Papin C, Rouyer F. The clockwork orange Drosophila protein functions as both an activator and a repressor of clock gene expression. J Biol Rhythms. 2008 Apr;23(2):103–16.

114. Meireles-Filho ACA, Bardet AF, Yáñez-Cuna JO, Stampfel G, Stark A. cis-regulatory requirements for tissue-specific programs of the circadian clock. Curr Biol. 2014 Jan 6;24(1):1–10.

115. Abruzzi KC, Rodriguez J, Menet JS, Desrochers J, Zadina A, Luo W, et al. Drosophila CLOCK target gene characterization: implications for circadian tissue-specific gene expression. Genes Dev. 2011 Nov 15;25(22):2374–86.

116. Menet JS, Pescatore S, Rosbash M. CLOCK:BMAL1 is a pioneer-like transcription factor. Genes Dev. 2014 Jan 1;28(1):8–13.

117. Glossop NRJ, Houl JH, Zheng H, Ng FS, Dudek SM, Hardin PE. VRILLE Feeds Back to Control Circadian Transcription of Clock in the Drosophila Circadian Oscillator. Neuron. 2003 Jan 23;37(2):249–61.

118. Wang T, Montell C. Phototransduction and retinal degeneration in Drosophila. Pflugers Arch - Eur J Physiol. 2007 Jun 21;454(5):821–47.

119. Allada R, Kadener S, Nandakumar N, Rosbash M. A recessive mutant of Drosophila Clock reveals a role in circadian rhythm amplitude. EMBO J. 2003 Jul 1;22(13):3367–75.

120. Vaccaro A, Issa AR, Seugnet L, Birman S, Klarsfeld A. Drosophila Clock Is Required in Brain Pacemaker Neurons to Prevent Premature Locomotor Aging Independently of Its Circadian Function. PLOS Genetics. 2017 Jan 10;13(1):e1006507.

121. B. Domènech E, Marfany G. The Relevance of Oxidative Stress in the Pathogenesis and Therapy of Retinal Dystrophies. Antioxidants (Basel) [Internet]. 2020 Apr 23 [cited 2021 Apr 6];9(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7222416/

122. ZITKA O, SKALICKOVA S, GUMULEC J, MASARIK M, ADAM V, HUBALEK J, et al. Redox status expressed as GSH:GSSG ratio as a marker for oxidative stress in paediatric tumour patients. Oncol Lett. 2012 Dec;4(6):1247–53.

123. Gorbatyuk M, Gorbatyuk O. Review: Retinal degeneration: Focus on the unfolded protein response. Mol Vis. 2013 Sep 20;19:1985–xxx.

124. Shinde V, Kotla P, Strang C, Gorbatyuk M. Unfolded protein response-induced dysregulation of calcium homeostasis promotes retinal degeneration in rat models of autosomal dominant retinitis pigmentosa. Cell Death Dis. 2016 Feb;7(2):e2085–e2085.

125. Geng C, Pak WL. Photoreceptor Degeneration and Ca2+ Influx through Light-Activated Channels of Drosophila [Internet]. Madame Curie Bioscience Database [Internet]. Landes Bioscience; 2013 [cited 2021 Sep 14]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6284/

126. Kuintzle RC, Chow ES, Westby TN, Gvakharia BO, Giebultowicz JM, Hendrix DA. Circadian deep sequencing reveals stress-response genes that adopt robust rhythmic expression during aging. Nature Communications. 2017 Feb 21;8(1):14529.

127. Damulewicz M, Świątek M, Łoboda A, Dulak J, Bilska B, Przewłocki R, et al. Daily Regulation of Phototransduction, Circadian Clock, DNA Repair, and Immune Gene Expression by Heme Oxygenase in the Retina of Drosophila. Genes (Basel). 2018 Dec 21;10(1):6.

128. Baba K, Piano I, Lyuboslavsky P, Chrenek MA, Sellers JT, Zhang S, et al. Removal of clock gene Bmal1 from the retina affects retinal development and accelerates cone photoreceptor degeneration during aging. PNAS. 2018 Dec 18;115(51):13099–104.

129. Sawant OB, Horton AM, Zucaro OF, Chan R, Bonilha VL, Samuels IS, et al. The Circadian Clock Gene Bmal1 Controls Thyroid Hormone-Mediated Spectral Identity and Cone Photoreceptor Function. Cell Rep. 2017 Oct 17;21(3):692–706.

130. Grewal R, Organisciak D, Wong P. Factors underlying circadian dependent susceptibility to light induced retinal damage. Adv Exp Med Biol. 2006;572:411–6.

131. Organisciak DT, Darrow RM, Barsalou L, Kutty RK, Wiggert B. Circadian-Dependent Retinal Light Damage in Rats. Investigative Ophthalmology & Visual Science. 2000 Nov 1;41(12):3694–701.

132. Kondratov RV, Kondratova AA, Gorbacheva VY, Vykhovanets OV, Antoch MP. Early aging and age-related pathologies in mice deficient in BMAL1, the core componentof the circadian clock. Genes Dev. 2006 Jul 15;20(14):1868–73.

133. Aging | Premature aging of the hippocampal neurogenic niche in adult *Bmal1*- deficient mice - Full Text [Internet]. [cited 2021 Apr 26]. Available from: https://www.aging-us.com/article/100764/text

134. Musiek ES, Lim MM, Yang G, Bauer AQ, Qi L, Lee Y, et al. Circadian clock proteins regulate neuronal redox homeostasis and neurodegeneration. J Clin Invest. 2013 Dec 2;123(12):5389–400.

135. Goyal V, DeVera C, Baba K, Sellers J, Chrenek MA, Iuvone PM, et al. Photoreceptor Degeneration in Homozygous Male Per2luc Mice During Aging. J Biol Rhythms. 2020 Nov 2;0748730420965285.

136. Yoo SH, Yamazaki S, Lowrey PL, Shimomura K, Ko CH, Buhr ED, et al. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A. 2004 Apr 13;101(15):5339–46.

137. Martínez-Águila A, Martín-Gil A, Carpena-Torres C, Pastrana C, Carracedo G. Influence of Circadian Rhythm in the Eye: Significance of Melatonin in Glaucoma. Biomolecules. 2021 Feb 24;11(3).

138. Ko GYP. Circadian regulation in the retina: From molecules to network. Eur J Neurosci. 2020 Jan;51(1):194–216.

139. Hodge BA, Meyerhof GT, Katewa SD, Lian T, Lau C, Bar S, et al. Dietary restriction and clock delay eye aging to extend lifespan in D. melanogaster [Internet]. 2021 May [cited 2021 Aug 29] p. 2021.05.08.443272. Available from: https://www.biorxiv.org/content/10.1101/2021.05.08.443272v1

140. Ma D, Li S, Molusky MM, Lin JD. Circadian autophagy rhythm: a link between clock and metabolism? Trends Endocrinol Metab. 2012 Jul;23(7):319–25.

141. Cavieres-Lepe J, Ewer J. Reciprocal Relationship Between Calcium Signaling and Circadian Clocks: Implications for Calcium Homeostasis, Clock Function, and Therapeutics. Frontiers in Molecular Neuroscience. 2021;14:83.

142. Mustafi D, Kevany BM, Genoud C, Bai X, Palczewski K. Photoreceptor phagocytosis is mediated by phosphoinositide signaling. FASEB J. 2013 Nov;27(11):4585–95.

143. Krishnan N, Davis AJ, Giebultowicz JM. Circadian regulation of response to oxidative stress in Drosophila melanogaster. Biochemical and Biophysical Research Communications. 2008 Sep 19;374(2):299–303.

144. Wilking M, Ndiaye M, Mukhtar H, Ahmad N. Circadian Rhythm Connections to Oxidative Stress: Implications for Human Health. Antioxid Redox Signal. 2013 Jul 10;19(2):192–208.

145. Rohowetz LJ, Kraus JG, Koulen P. Reactive Oxygen Species-Mediated Damage of Retinal Neurons: Drug Development Targets for Therapies of Chronic Neurodegeneration of the Retina. Int J Mol Sci. 2018 Oct 27;19(11):3362.

146. Jauregui-Lozano J, Easton A, Escobedo S, Lanman NA, Weake VM, Hall H. Proper control of R-loop homeostasis is required for maintenance of gene expression and neuronal function during aging [Internet]. Molecular Biology; 2021 Jun [cited 2021 Jul 13]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.06.29.450380

147. Ferreiro MJ, Pérez C, Marchesano M, Ruiz S, Caputi A, Aguilera P, et al. Drosophila melanogaster White Mutant w1118 Undergo Retinal Degeneration. Frontiers in Neuroscience. 2018;11:732.

148. Lugena AB, Zhang Y, Menet JS, Merlin C. Genome-wide discovery of the daily transcriptome, DNA regulatory elements and transcription factor occupancy in the monarch butterfly brain. PLoS Genet [Internet]. 2019 Jul 23 [cited 2021 Apr 11];15(7). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6677324/

149. Luo W, Chen WF, Yue Z, Chen D, Sowcik M, Sehgal A, et al. Old flies have a robust central oscillator but weaker behavioral rhythms that can be improved by genetic and environmental manipulations. Aging Cell. 2012;11(3):428–38.

150. Rakshit K, Krishnan N, Guzik EM, Elżbieta Pyza, Giebultowicz JM. Effects of Aging on the Molecular Circadian Oscillations in Drosophila. Chronobiology International. 2012 Feb 1;29(1):5–14.

151. Litovchenko M, Meireles-Filho ACA, Frochaux MV, Bevers RPJ, Prunotto A, Anduaga AM, et al. Extensive tissue-specific expression variation and novel regulators underlying circadian behavior. Sci Adv. 2021 Jan 29;7(5):eabc3781.

152. Ulgherait M, Chen A, McAllister SF, Kim HX, Delventhal R, Wayne CR, et al. Circadian regulation of mitochondrial uncoupling and lifespan. Nat Commun. 2020 Apr 21;11(1):1927.

153. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

154. Ambrosini G, Groux R, Bucher P. PWMScan: a fast tool for scanning entire genomes with a position-specific weight matrix. Bioinformatics. 2018 Jul 15;34(14):2483–4.

155. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002 Jun 1;12(6):996–1006.

156. Vera Alvarez R, Pongor LS, Mariño-Ramírez L, Landsman D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. Bioinformatics. 2019 Jun 1;35(11):1960–2.

157. Chen X, Hall H, Simpson JP, Leon-Salas WD, Ready DF, Weake VM. Cytochrome b5 protects photoreceptors from light stress-induced lipid peroxidation and retinal degeneration. NPJ Aging Mech Dis. 2017 Dec 4;3:18.

158. Florholmen-Kjær Å, Lyså RA, Fuskevåg OM, Goll R, Revhaug A, Mortensen KE. A sensitive method for the analysis of glutathione in porcine hepatocytes. Scand J Gastroenterol. 2014 Nov;49(11):1359–66.

159. Jackson GR, Owsley C, Curcio CA. Photoreceptor degeneration and dysfunction in aging and age-related maculopathy. Ageing Research Reviews. 2002 Jun 1;1(3):381–96.

160. Aguilera A, García-Muse T. R Loops: From Transcription Byproducts to Threats to Genome Stability. Molecular Cell. 2012 Apr 27;46(2):115–24.

161. Niehrs C, Luke B. Regulatory R-loops as facilitators of gene expression and genome stability. Nature Reviews Molecular Cell Biology. 2020 Jan 31;1–12.

162. Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. PNAS. 2015 May 5;112(18):5785–90.

163. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Molecular Cell. 2011 Jun 24;42(6):794–805.

164. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature. 2014 Dec;516(7531):436–9.

165. Mackay RP, Xu Q, Weinberger PM. R-Loop Physiology and Pathology: A Brief Review. DNA and Cell Biology. 2020 Oct 14;39(11):1914–25.

166. McKinnon PJ. Topoisomerases and the regulation of neural function. Nat Rev Neurosci. 2016 Nov;17(11):673–9.

167. Kwan KY, Wang JC. Mice lacking DNA topoisomerase IIIβ develop to maturity but show a reduced mean lifespan. Proceedings of the National Academy of Sciences. 2001;98(10):5717–21.

168. Yan Q, Sarma K. MapR: A Method for Identifying Native R-Loops Genome Wide. Current Protocols in Molecular Biology. 2020;130(1):e113.

169. Rauser CL, Abdel-Aal Y, Shieh JA, Suen CW, Mueller LD, Rose MR. Lifelong heterogeneity in fecundity is insufficient to explain late-life fecundity plateaus in Drosophila melanogaster. Experimental Gerontology. 2005 Aug 1;40(8):660–70.

170. Hall H, Medina P, Cooper DA, Escobedo SE, Rounds J, Brennan KJ, et al. Transcriptome profiling of aging Drosophila photoreceptors reveals gene expression trends that correlate with visual senescence. BMC Genomics. 2017 Nov 21;18(1):894.

171. Alecki C, Chiwara V, Sanz LA, Grau D, Arias Pérez O, Boulier EL, et al. RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. Nat Commun. 2020 Apr 14;11(1):1781.

172. Belotserkovskii BP, Tornaletti S, D'Souza AD, Hanawalt PC. R-loop generation during transcription: Formation, processing and cellular outcomes. DNA Repair. 2018 Nov 1;71:69–81.

173. Allison DF, Wang GG. R-loops: formation, function, and relevance to cell stress. Cell Stress. 2019 Jan 21;3(2):38–46.

174. Chedin F, Benham CJ. Emerging roles for R-loop structures in the management of topological stress. J Biol Chem. 2020 Apr 3;295(14):4684–95.

175. Lopes I, Altab G, Raina P, de Magalhães JP. Gene Size Matters: An Analysis of Gene Length in the Human Genome. Frontiers in Genetics. 2021;12:30.

176. Stoeger T, Grant RA, McQuattie-Pimentel AC, Anekalla K, Liu SS, Tejedor-Navarro H, et al. Aging is associated with a systemic length-driven transcriptome imbalance. bioRxiv. 2019 Jul 3;691154.

177. Tous C, Aguilera A. Impairment of transcription elongation by R-loops in vitro. Biochemical and Biophysical Research Communications. 2007 Aug 24;360(2):428–32.

178. Bentin T, Cherny D, Larsen HJ, Nielsen PE. Transcription arrest caused by long nascent RNA chains. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 2005 Feb 14;1727(2):97–105.

179. King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, et al. Topoisomerases facilitate transcription of long genes linked to autism. Nature. 2013 online;501:58.

180. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. PNAS. 1987 Oct 1;84(20):7024–7.

181. Pommier Y, Sun Y, Huang S yin N, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nature Reviews Molecular Cell Biology. 2016 online;17:703.

182. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. Nature Reviews Molecular Cell Biology. 2002 Jun;3(6):430–40.

183. Xu D, Shen W, Guo R, Xue Y, Peng W, Sima J, et al. Top3β is an RNA topoisomerase that works with Fragile X syndrome protein to promote synapse formation. Nat Neurosci. 2013 Sep;16(9):1238–47.

184. Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. Arginine Methylation Facilitates the Recruitment of TOP3B to Chromatin to Prevent R Loop Accumulation. Molecular Cell. 2014 Feb 6;53(3):484–97.

185. Joo Y, Xue Y, Wang Y, McDevitt RA, Sah N, Bossi S, et al. Topoisomerase 3β knockout mice show transcriptional and behavioural impairments associated with neurogenesis and synaptic plasticity. Nature Communications. 2020 Jun 19;11(1):3143.

186. Hall H, Cooper BR, Qi G, Wijeratne AB, Mosley AL, Weake VM. Quantitative Proteomic and Metabolomic Profiling Reveals Altered Mitochondrial Metabolism and Folate Biosynthesis Pathways in the Aging Drosophila Eye. Mol Cell Proteomics. 2021 Jul 29;20:100127.

187. Choe K, Clandinin TR. Thinking about Visual Behavior; Learning about Photoreceptor Function. In: Current Topics in Developmental Biology [Internet]. Academic Press; 2005 [cited 2021 Jun 18]. p. 187–213. (Neural Development; vol. 69). Available from: https://www.sciencedirect.com/science/article/pii/S0070215305690072

188. Simon AF, Liang DT, Krantz DE. Differential decline in behavioral performance of Drosophila melanogaster with age. Mechanisms of Ageing and Development. 2006 Jul 1;127(7):647–51.

189. Carbone MA, Yamamoto A, Huang W, Lyman RA, Meadors TB, Yamamoto R, et al. Genetic architecture of natural variation in visual senescence in Drosophila. PNAS. 2016 Oct 25;113(43):E6620–9.

190. Grotewiel MS, Martin I, Bhandari P, Cook-Wiens E. Functional senescence in Drosophila melanogaster. Ageing Research Reviews. 2005 Aug 1;4(3):372–97.

191. Zylka MJ, Simon JM, Philpot BD. Gene Length Matters in Neurons. Neuron. 2015 Apr 22;86(2):353–5.

192. Stegeman R, Hall H, Escobedo SE, Chang HC, Weake VM. Proper splicing contributes to visual function in the aging Drosophila eye. Aging Cell. 2018;17(5):e12817.

193. Groh M, Gromak N. Out of Balance: R-loops in Human Disease. PLOS Genetics. 2014;10(9):e1004630.

194. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, et al. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Molecular Cell. 2016 Jul 7;63(1):167–78.

195. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev. 2016 Jun 1;30(11):1327–38.

196. Zhang T, Wallis M, Petrovic V, Challis J, Kalitsis P, Hudson DF. Loss of TOP3B leads to increased R-loop formation and genome instability. Open Biology. 2020 Jan 9;9(12):190222.

197. Hu Y, Chung V, Comjean A, Rodiger J, Nipun F, Perrimon N, et al. BioLitMine: Advanced Mining of Biomedical and Biological Literature About Human Genes and Genes from Major Model Organisms. G3 (Bethesda). 2020 Dec 1;10(12):4531–9.

VITA

Juan Jauregui-Lozano

General Information

Nationality: Colombian

Birth date: March 28th, 1993

Education

August 2017 – May 2022	Purdue University, PhD in Biochemistry
January 2011 - May 2017	National University of Colombia,

B.S. in Biology

Research Experience

March 2018 – May 2022   Purdue University, Biochemistry department

Graduate Research Assistant.

Research: Elucidating the mechanisms that contribute to the aging transcriptome of Drosophila photoreceptors.

Directed by Dr. Vikki M. Weake

June 2016 - December 2016   Purdue University, Biochemistry department

Undergraduate Research Assistant – Part of International Undergraduate Research Experience.

Research: Identification of kinase substrates using bacterial expression systems.

Directed by Dr. W. Andy Tao.

January 2015 - March 2016   Purdue University

Undergraduate Research Assistant.

Research: Identification of essential enzymes associated with NAD metabolism in *T. vaginalis*.

Directed by Dr. rer. Nat. Maria H. Ramirez

Publications

1) In vivo tissue-specific chromatin profiling in Drosophila melanogaster using GFP-tagged nuclei. Jauregui-Lozano, J., Bakhle, K. & Weake, V.M.

Genetics, Volume 218, Issue 3, July 2021, iyab079. https://doi.org/10.1093/genetics/iyab079

2) Proper control of R-loop homeostasis is required for maintenance of gene expression and neuronal function during aging.

Jauregui-Lozano, J., Escobedo, E., Easton, A., Weake, V.M. & Hall, H.

Aging Cell. 2022 Jan 20:e13554. doi: 10.1111/acel.13554. https://doi.org/10.1111/acel.13554

3) The Clock:cycle complex is a major transcriptional regulator of Drosophila photoreceptors that protects the eye from retinal degeneration and oxidative stress.

Jauregui-Lozano, J., Hall, H., Stanhope, S., Bakhle, K., Marlin, M., & Weake, V.M.

PLOS Genetics (Accepted 1/8/22). Also on biorXiv https://doi.org/10.1101/2021.09.27.461946

Honors/funding/awards

Spring 2021   Bird Stair Research Fellowship

Biochemistry department – Purdue $3000

Research grant for project: "Identifying splicing regulators of the aging clock in Drosophila photoreceptor neurons"

Summer and Fall 2021   Ross-Lynn Research Scholar fund

Biochemistry department – Purdue $31000

7-month award that provides a salary. Based on excellence in research, including quantity and quality of published work, and the quality of scholarship demonstrated in the written application.

Spring 2019   Bird Stair Research Fellowship

Biochemistry department – Purdue $5000

Research grant for project: "Active histone marks contribute to photoreceptor health during aging by regulating gene expression"

June 2016 – December 2016   Travel grant for international undergraduate research
experience
National University of Colombia

Presentations

Jauregui-Lozano, J., Weake, V. M. Tissue-specific chromatin profiling reveals a key role for
Clock-dependent transcription in regulation of Drosophila photoreceptor homeostasis. Oral
presentation at the 63rd Annual Drosophila Research Conference at San Diego CA, April 2022

Jauregui-Lozano, J., Weake, V. M. Integration of ATAC-seq and RNA-seq in aging
photoreceptors reveals a role for Clock:cycle in regulation of phototransduction gene expression
and retinal degeneration. Poster presented at 2021 Course "Impacts of Sleep and Circadian
Biology on Alzheimer's disease and Aging" at the Jackson Laboratory, ME

Jauregui-Lozano, J., Weake, V. M. Optimization of chromatin profiling techniques using tissue-
specific samples. Poster presented at 2019 Course "Attacking Open Chromatin with ATAC
sequencing", at the European Molecular Biology Laboratory, Germany.

Training and workshops

May 2020   Coursera (Virtual)
Programming for everybody (Getting started with Python)
August 26th-August 30th 2019   EMBL, Germany
Attacking Open Chromatin with ATAC sequencing
October 24th-October 28th 2021   The Jackson Laboratory, Bar Harbor ME
Impacts of Sleep and Circadian Biology on Alzheimer's disease and Aging

Teaching Experience

August 2020- December 2020   Purdue University
Teaching assistant
August 2018- December 2018   Purdue University
Teaching assistant
January 2017- March 2017   National University of Colombia
Teaching assistant coordinator

Sept. 2015- January 2016   National University of Colombia

Teaching assistant

January 2015- July 2015   National University of Colombia

Teaching assistant

Mentoring

Kimaya Bakhle – Biochemistry department	Undergraduate student	Spring 2019-Spring 2021
Arriana Higgins	NSF REU student	Summer 2019
Makayla Marlin – Biochemistry department	Undergraduate student	Spring 2021 – present
Grace Meng – Biochemistry department	PhD rotation student	Fall 2021

Extracurricular, community and avocational activities

March 20th, 2018   Wea Ridge Elementary School, Lafayette IN

Volunteer at "Science in Schools" Night Event

January 2012-December 2012   National University of Colombia, Colombia

Volunteer at Wild Fauna Centre for Rescue and Rehabilitation (URRAS)