# AMPK counteracts anomalous R-loops at open promoters in starved C. elegans germlines

## Reference

Sun, B., Sherrin, M., Roy, R., 2023. Unscheduled epigenetic modifications cause genome instability and sterility through aberrant R-loops following starvation. *Nucleic Acids Research*. 51:1. 84-98.

## Introduction

Chromatin dynamics are at the heart of successful meiosis and proper regulation of epigenetic modifications associated with transcriptional processes is essential to proper chromosome segregation during meiotic divisions. While DNA double strand breaks (DSBs) are abundant and, contrarily to cycling nuclei, systematically required in meiotic nuclei, their formation is a tightly-regulated and programmed process. DSB formation favors open chromatin([Price & Andrea, 2013](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3670600/)) and unscheduled euchromatic regions can therefore poteniate unprogrammed DSBs in turn. In meiotic nuclei, failure to repair aberrant DSBs is a significant threat to gamete genome integrity and defects can manifest transgenerationally or lead to reproductive sterility as a result. Therefore, understanding the nuclear chromatin environment and the regulatory network surrounding chromatin dynamics is of major interest to meiosis research and is key to capturing molecular targets in clinical applications. When a system is stressed, the epigenetic environment undergoes significant changes in response as gene expression requirements are shifted ([Gudsnuk & Champagne, 2012](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4021821/)). [Sun, Sherrin, & Roy, 2023](https://academic.oup.com/nar/article/51/1/84/68876020) found that under starvation conditions, H3K4me3 marks are significantly elevated in *Caenorhabditis elegans* meiotic nuclei at open promoters in accompaniment with aberrant R-loops when the AMP kinase is absent. They propose that R-loops accumulate as the result of hyperactive transcription during starvation and then hijack key repair proteins from programmed DSB sites, leading to aberrant repair and transgenerational sterility. However, studies performed in yeast models ([Yang et. al, 2022](https://www.sciencedirect.com/science/article/pii/S2211124721015916)) suggest that R-loops may actually drive DSB repair in meiosis under wild-type conditions. While a field consensus on R-loops’ full contribution to the regulation of meiosis has not yet been reached, it is clear that they are associated with meiotic DSBs in some capacity, and therefore may aid in the identification of DSB hotspots. My aim is thus to compile a picture of where R-loop reads map on the whole-genome scale in an effort to identify candidate hotspot regions.

## Expected Figure

The expected figure would portray information from [Sun, Sherrin, & Roy, 2023](https://www.sciencedirect.com/science/article/pii/S2211124721015916) (Figure 3). A major motivation for this project however is that the authors were interrogating the contribution of H3K4me3 to the chromatin landscape and found that they had accrued R-loops, but only focused on sites that overlapped with sites that also had H3K4me3. My interest lies more in the effects of the R-loops themselves, and I am therefore aiming to generate a figure that reflects *all* of their reads (as in example heat map below).

**Figure 3: R-loop formation correlates with ectopic deposition of H3K4me3**

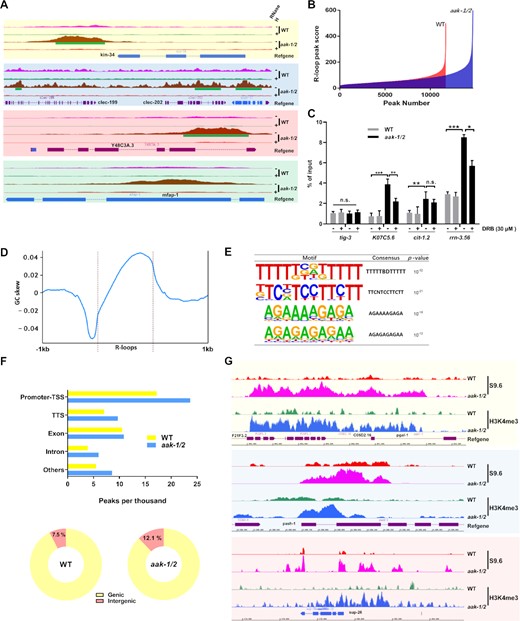
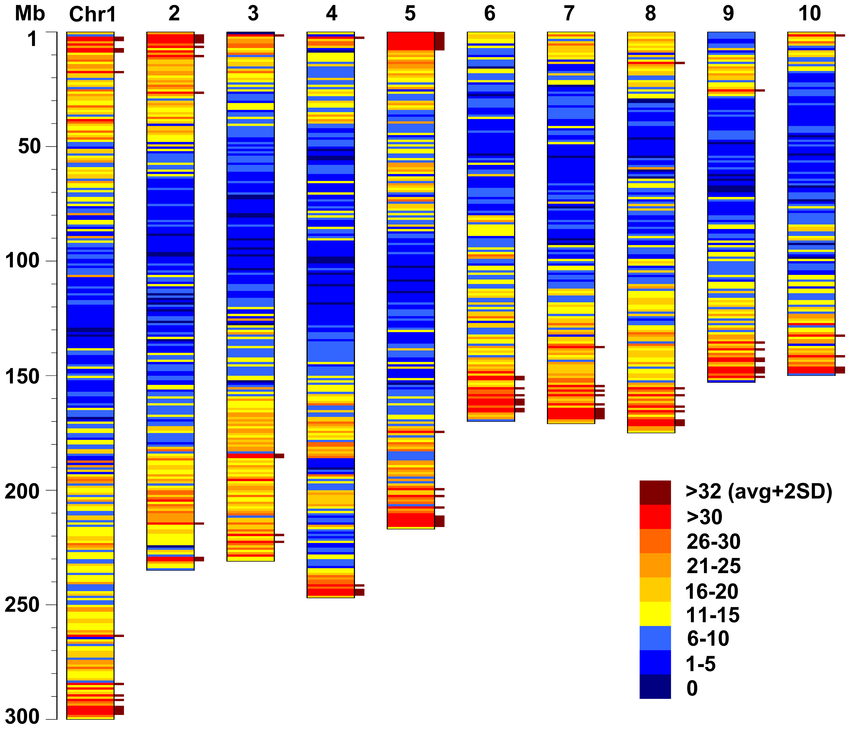


Figure 3: R-loop formation correlates with ectopic deposition of H3K4me3

**Legend from** [**Sun, Sherrin, and Roy, 2023**](https://www.sciencedirect.com/science/article/pii/S2211124721015916) (A) Genome browser snapshots of DRIP-seq signals at regions proximal to genes and RNH tracks. Green bars show R-loop peak calls. Track height represents read counts. (B) Overall comparison of DRIP-seq output (peak score and number) in WT versus *aak-1/2* genome. (C) DRIP-qPCR validation. WT or F2 descendants of starved *aak-1/2* mutants cultured with or without DRB treatment during starvation for 3 days were collected for DRIP. *tig-3* was selected as a negative control; n = 3, mean ± S.E.M. Signal values normalized with respect to input genomic DNA are plotted. *P< 0.05,* ***P< 0.01,*** P < 0.001 by Mann–Whitney U-test. (D) Metaplot of GC skew centered on all R-loop peaks. (E) Four enriched *de novo* motifs identified by HOMER analysis of AMPK mutant versus WT DRIP-seq. (F) Upper: annotation and peak location analyses for called R-loop peaks mapping to promoter–TSS, TTS, exon, intron and the other regions. Genomic peak proportion in per thousand is shown. Peaks accumulate predominantly at the promoter–TSS fraction in *aak-1/2* mutants. Lower: pie charts of DRIP-seq signal distribution for genic versus intergenic regions in WT or F2 descendants of starved aak-1/2 mutants. (G) Genome browser snapshots showing a representative sample of positively correlated R-loop signals and H3K4me3 levels in the F2 descendants of starved AMPK mutants.

This figure is showing the results of the authors’ DNA-RNA immunprecipitation followed by high throughput sequencing (DRIP-Seq) experiments. This method captures R-loops by antibody immunoprecipitation and the immunoprecipitated products are then sequenced to identify regions of the genome at which R-loops were associated in the sample. In panel A, a genome browser snapshot is showing some example genes where reads aligned and is validating that the signal is truly from R-loops (RNaseH is an enzyme that removes R-loops, see control). Panel B is comparing the number of peak reads which came from AMPK mutants (*aak-1/2*) as compared to wild type (WT). Panel C also uses the DRIP technique, however instead of sequencing the immunoprecipitate, a quantitative PCR was run for the same representative genes as shown in panel A. Panels D-F represent R-loop enrichment at various sequence features. Panel G is showing the overlap of reads from their DRIP-seq experiments at these representative genes with those of a similar experiment performed in which they immunoprecipitated and sequenced H3K4me3 (S9.6 is an antibody against R-loops).

**Example Heat Map from** [**StackOverflow**](https://i.stack.imgur.com/EpF3I.png)



Example Heat Map

Representative image of expected final project output; note that *C. elegans* have only five autosomes and one X chromosome to plot, contrary to what is pictured. Legend would reflect thousands of reads.

## Materials and Methods

The authors made their sequencing data publicly-available in a repository with NCBI. I have already downloaded these reads and the Bioinformatics core at IIHG recently provided an Excel spreadsheet to me with the processed reads. Further, my undergraduate research assistant just finished annotating this file. At this time, I have the data available to me with genomic coordinates and the number of reads which mapped to each coordinate from both the authors’ AMPK mutants and their wild-type samples, as well as the total input DNA from both backgrounds. From here, I need to graph the figure in R by chromosome, however given I have never used R before, I imagine this will be more of an undertaking than it sounds.

The example image I have provided from StackOverflow unfortunately comes from a thread that does not provide the original citation for the image. However, the thread is regarding the generation of a heat map by chromosome in R, which is what I am trying to do with my dataset, and the thread has some helpful suggestions StackOverflow thread The first suggestion is to use ggplot, which would generate a red, yellow, and blue heat map oriented very similarly to the example image:

‘ggplot(dt) + geom\_rect(aes(xmin=chromosome - 0.3, xmax=chromosome + 0.3, ymin=mb\_from, ymax=mb\_to, fill=score)) + scale\_y\_reverse() + scale\_fill\_distiller(palette=“RdYlBu”)’

An alternative suggestion in this thread is to use the R package “plotly” and would create a plot which places the chromosome number on the x-axis and the chromosomal position on the y-axis (also very similar to the example image, although this approach will make the graphical representation messier as it does not include space between individual chromosomes):

‘dat <- apply(dt, 1, function(x) data.table(chromosome = x[“chromosome”], mb = x[“mb\_from”]:x[“mb\_to”], score = x[“score”]) ) %>% rbindlist() plot\_ly(dat, x = ~chromosome, y = ~mb, z = ~score, type = “heatmap”, colors = “RdYlBu”, reversescale = T) %>% layout(yaxis = list(range = c(1000, 0)))’

Because I have heard of ggplot and am completely unfamiliar with R in general, I feel more confident using the ggplot approach to start, however should that fail, I will then revert to trying the plotly package or exploring other avenues.

## Reflections

In reviewing the StackOverflow thread, I am concerned that I need to rearrange my data formatting, however I’m not clear on the format that the original poster used in their dataset that they are receiving these suggestions for. My next step will be to determine what precisely their data layout looks like (it is presented as R code and I just don’t get it yet) and reformat my own data to match that. I know that the dataset I have has all the information required, it is just a matter of reorganizing column headers to fit the suggested scripts. I am also totally unsure why my first figure is coming out so small when converted through pandoc for submission in the main README for the project, but I have tried inserting more spaces around the image, using both a web-based link and a local image file, and using pandoc to convert to .docx instead of .pdf and none of these attempts have resolved the issue. For a full image size, please refer to this README file in my GitHub repository.