**RerepSeq Data Analysis and Figure Source Code**

**Overview**

All code used to perform the analysis and generate the figures for our genome-wide sequencing technique, Rerep-Seq, is freely available and free to use unconditionally at <https://github.com/blacklabUCD/ReRepSeqMethods>. This code relies on basic uses of R (version 3.6) to normalize and process bedgraph files, as well as basic implementations of existing command-line software like sqlite3. For convenience, we have organized our code repository by figure and supplied detailed descriptions of the source code functionality in individual readme files for each program as well as in-line comments within the code. After installing the dependencies, and with a basic familiarity of the R environment, users will be able to generate the figures presented in the manuscript.

The following is an overview of the contents of the ReRepSeqMethods code repository:

* Blacklist files- bed files denoting regions of the yeast and human genome with high background that were excluded from the analyses.
* Replication timing files- bed files from published data denoting regions of the genome as early or late replicating domains or identifying ARS elements in yeast.
* Data normalization, scaling, binning, and smoothing
* Code by figures
  + Figure 3
    - Panel A
    - Panel B
    - Panel C
  + Figure 5
    - Panel A
    - Panel B
    - Panel C
  + Figure 6
    - Panel E
  + Supplemental Figure 1
    - Panel A
    - Panel B
  + Supplemental Figure 2
    - Panel A
    - Panel B
  + Supplemental Figure 3
    - Panel A
    - Panel B

**System Requirements**

The source code in the ReRepSeqMethods repository was written and executed using the 64bit version of Ubuntu 18.04 operating system, and has been successfully executed on a laptop with an Intel® Core™ i7-4600M CPU @ 2.90GHz × 4, 15.5 GiB of ram, and 2TB of disk space. While a system of this size is not necessarily required, smaller machines have not been tested. Fastq files were aligned using a larger machine, also running a 64bit version of Ubuntu 18.04. This machine featured an Intel® Xenon™ processor CPU @ 3.01GHz x 24, 128 GiB of ram, and 48TB of disk space.

The expected run times for each piece of software available on our github repository is less than 15 minutes using an equivalent system.

**Dependencies**

R packages: [dplyr](https://cran.r-project.org/web/packages/dplyr/index.html), [readr](https://cran.r-project.org/web/packages/readr/index.html), [tidyr](https://cran.r-project.org/web/packages/tidyr/index.html), [ggplot2](https://cran.r-project.org/web/packages/ggplot2/index.html), [rtracklayer](https://bioconductor.org/packages/release/bioc/html/rtracklayer.html), [GenomicRanges](https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html), [ggpubr](https://cran.r-project.org/web/packages/ggpubr/index.html), and [reshape2](https://cran.r-project.org/web/packages/reshape2/index.html).

Linux command line: [sqlite3](https://www.sqlite.org/index.html)

**Key Operations**

The source code will take users from an un-normalized bedgraph outputted by bedtools, to a normalized, scaled, and final bedgraph that will be used to generate all figures in our manuscript. The code to generate each figure in the manuscript is also supplied and accompanied with the necessary input files to test the provided code. Detailed descriptions for each program are provided in a separate readme file located in the same subdirectory as the program and input files.

**Key Characteristics**

To run the programs provided in the ReRepSeqMethods repository, users will need to install R version 3.6. This is necessary due to the use of [Gviz](https://bioconductor.org/packages/release/bioc/html/Gviz.html), version 1.3, to generate the whole chromosome track figures featured in figures 3a, 5a, 6e, and supplemental figures 1a, 2a, and 3a. Additionally, users will need to install the R packages: [dplyr](https://cran.r-project.org/web/packages/dplyr/index.html), [readr](https://cran.r-project.org/web/packages/readr/index.html), [tidyr](https://cran.r-project.org/web/packages/tidyr/index.html), [ggplot2](https://cran.r-project.org/web/packages/ggplot2/index.html), [rtracklayer](https://bioconductor.org/packages/release/bioc/html/rtracklayer.html), [GenomicRanges](https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html), [ggpubr](https://cran.r-project.org/web/packages/ggpubr/index.html), and [reshape2](https://cran.r-project.org/web/packages/reshape2/index.html). Finally, users will also need to install the command line program [sqlite3](https://www.sqlite.org/index.html), which is used to bin the data.

**Test Data**

We have provided the data necessary to replicate for each figure in the manuscript. These files are located in the directory named after the figure they produce.