Natural variation in R-loop formation in *Drosophila melanogaster*

Timothy J. Stanek, Rohan M. Mehra, Christopher E. Ellison

Department of Genetics, Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ 08835

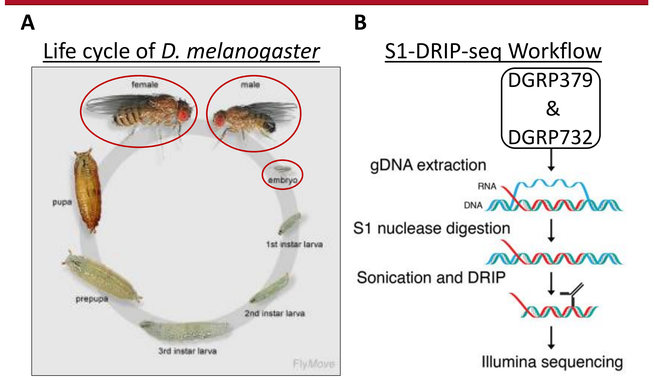
Program #867C

Introduction

R-loops are three-stranded nucleotide structures consisting of a DNA:RNA hybrid and a displaced ssDNA non-template strand. Originally viewed as a byproduct of transcription, R-loops are now recognized as important regulators of gene expression and genomic stability. R-loops are primarily associated with active Pol II transcription but have also been found at tRNA genes, mitochondrial genes, and transposable elements. Persistent dysregulation of R-loop maintenance can result in replication stress, DNA double-strand breaks, and chromosomal rearrangements that contribute to diseases such as neurological disorders and cancer. Although R-loops are conserved across cell types in mammals, little is known about natural variation in R-loop formation between individuals and throughout development.

Using DNA:RNA immunoprecipitation followed by high-throughput sequencing (DRIP-seq), we have mapped the R-loop profiles of two *D. melanogaster* individuals from the Drosophila Genetic Reference Panel (DGRP) at both embryonic and adult stages. Our current work assesses *in vivo* the underlying sequence and architectural determinants of R-loop formation and variation with respect to development and sex.

Methods and Materials



A) Embryos, adult females, and adult males were collected and used for subsequent DRIP-seq. B) Two wild strains of *D. melanogaster* DGRP379 and DGRP732 were processed for DRIP-seq from the stages indicated in *A*. S1 nuclease digestion preserves R-loop structures during sonication, allowing for immunoprecipitation of R-loops from highly repetitive regions of the genome.

DRIP-seq peak-calling and shared peaks overlap

Table 1. DRIP-seq peak-calling and shared peaks across conditions

	Embryos		Females		Males	
	DGRP-379	DGRP-732	DGRP-379	DGRP-732	DGRP-379	DGRP-732
Rep1	8316	8644	20783	19438	14287	15952
Rep2	10467	6916	18194	19340	13636	15429
Shared	*4797*	*3718*	*14056*	*14668*	*9933*	*10943*

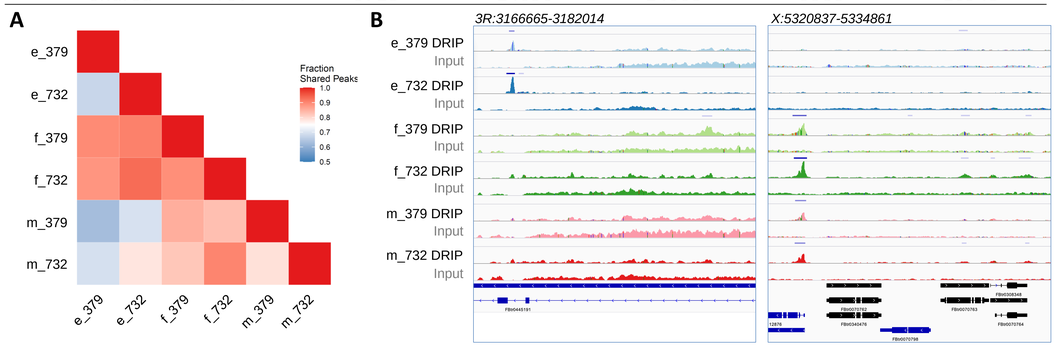


Table 1. Aligned DRIP-seq reads were used to call peaks with MACS2 (p-value cutoff = $1x10^{-3}$). Shared peaks between replicates were determined using *bedtools intersect*. A) Cross-conditional peak analysis plotted as fraction of shared peaks relative to the mate with fewer peaks called. B) DRIP-seq peaks unique to embryos (left) or adults (right).

Irreproducible Discovery Rate (IDR) Analysis

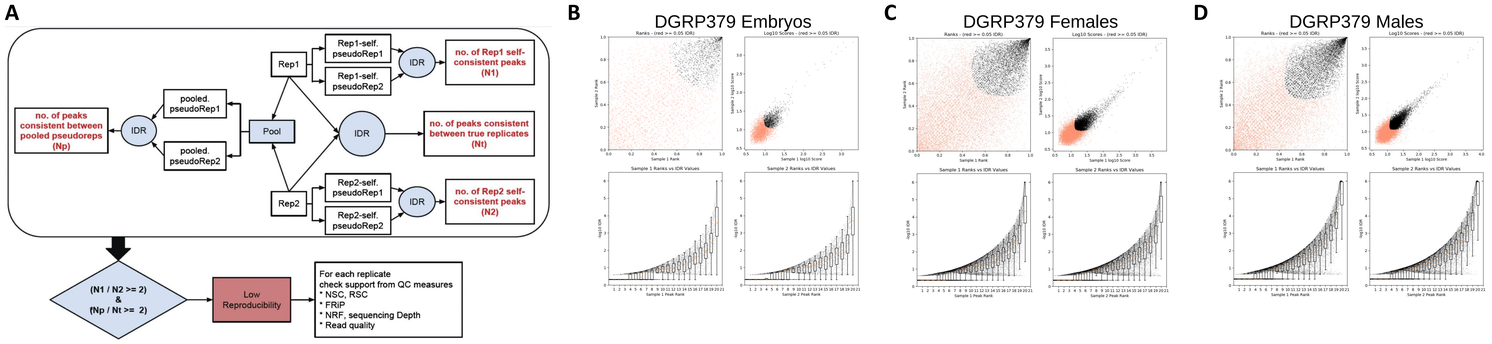
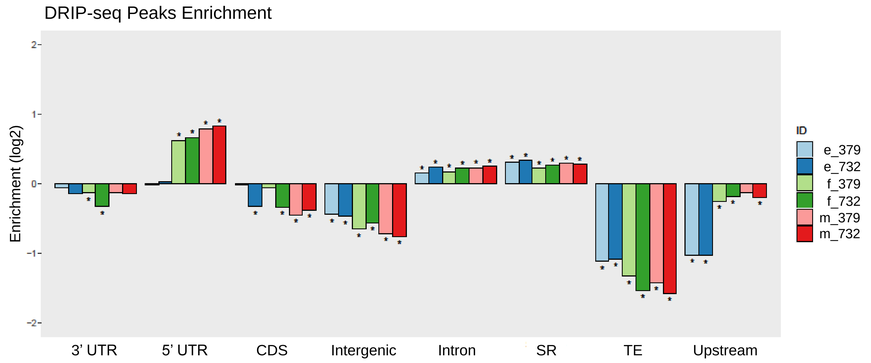


Table 2. IDR analysis determines the rate of reproducibility between replicates.

	Embryos		Females		Males	
	DGRP-379	DGRP-732	DGRP-379	DGRP-732	DGRP-379	DGRP-732
Nt	1352	936	5156	6170	3763	4227
Np	2778	1982	8073	8091	5836	6927
Np/Nt	*2.05*	*2.12*	*1.57*	*1.31*	*1.55*	*1.64*
N1	619	648	3824	3374	2639	3051
N2	793	389	2460	3094	1880	3175
N1/N2	*1.28*	*1.03*	*1.22*	*1.03*	*1.13*	*1.04*

A) IDR workflow from *ENCODE ChIP-seq Guidelines*. B-D) IDR Analysis of DGRP379 for embryos (B), females (C), and males (D). For each plot: *Upper Left* = Rep1 vs Rep2 peak ranks; peaks with IDR < 0.05 are colored red. *Upper right* = Rep1 vs Rep2 log10 peak scores; peaks with IDR < 0.05 are colored red. *Bo bottom row* = peak ranks vs IDR score for each true replicate. Table 2) Total peaks with IDR < 0.05 between true replicates (Nt) and pooled pseudo-replicates (Np) as well as between self-pseudoreplicates (N1 and N2) below a factor of 2 indicate good reproducibility.

R-loops are enriched/depleted at specific genetic elements



DRIP-seq peak enrichment (observed/expected) at specific genetic features was determined using *bedtools intersect*: 5' UTR, 3' UTR, CDS, Intergenic, Introns, short repeats, transposable elements (TE) and upstream regions. * indicates $p < 0.05$.

Acknowledgments and References

We thank the Genomics Facility at the Waksman Institute of Microbiology for use of the Covaris S220 sonicator and Agilent BioAnalyzer. We also acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here.

1. 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;30(11):1327–1338. doi:10.1101/gad.280834.116
2. Landt SG, Marinov GK, Kundaje A, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813–1831. doi:10.1101/gr.136184.111