nature research

Corresponding author(s): Julius Brennecke

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were chosen as large as possible while still practically feasible in terms of data collection. Adequate statistics has been applied throughout the manuscript in order to make sure that the observed effects are significant given the reported sample size.

2. Data exclusions

Describe any data exclusions.

no data exclusion in this manuscript.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful, except when caused by technical issues (e.g. material loss during handling leading to low-complexity NGS libraries)

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not relevant as grouping was not applied.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No investigator blinding was applied during data acquisition or analyses as the data was mostly analyzed in bulk by (blind) scripts such as for NGS or FISH quantification or blinding was not desirable for data presentation (e.g. blinded loading of western blots).

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We provide a link to a GitHub depository for the custom code that was used in this manuscript.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	there are no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Antibodies used in this study have either been validated in previous publications (cited in this manuscript) or have been generated for this particular manuscript (anti-Deadlock). Validation of this antibody was by using Deadlock mutants and observing loss of the signal in immuno-fluorescence experiments.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	used cell line in this study: Drosophila melanogaster Schneider cells. This is a standard cell line used in the field.
b. Describe the method of cell line authentication used.	does not apply
c. Report whether the cell lines were tested for mycoplasma contamination.	cells are routinely controlled for mycoplasm infection in the in house facility.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	does not apply

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	This study involved exclusively work with Drosophila melanogaster, a standard invertebrate model organism that does not underlie any ethical restrictions. Standard laboratory procedures have been applied throughout the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	does not apply

nature research

Corresponding Author: Julius Brennecke

Date: 2017-06-07

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. (The entry may remain private before publication.)

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=yvcbcqeybvylfkp&acc=GSE97719

3. Provide a list of all files available in the database submission.

ChIPseq_antiRhino_w1118_1_22682_uniq.bw
ChIPseq_antiRhino_Rhino_rhino_KO_22683_uniq.bw
ChIPseq_antiRhino_Rhino_moonshiner_KO_22684_uniq.bw
ChIPseq_antiRhino_Pld_promoter_deletion_43186_uniq.bw
ChIPseq_antiRNAPol2_w1118_1_46327_uniq.bw
ChIPseq_antiRNAPol2_rhino_KO_46330_uniq.bw
ChIPseq_antiRNAPol2_moonshiner_KO_46333_uniq.bw
ChIPseq_input_w1118_1_46326_uniq.bw
ChIPseq_input_rhino_KO_46329_uniq.bw
ChIPseq_input_moonshiner_KO_46332_uniq.bw
ChIPseq_antiRhino_w1118_1_22682.bam
ChIPseq_antiRhino_Rhino_rhino_KO_22683.bam
ChIPseq_antiRhino_Rhino_moonshiner_KO_22684.bam
ChIPseq_antiRhino_Pld_promoter_deletion_43186.bam
ChIPseq_antiRNAPol2_w1118_1_46327.bam
ChIPseq_antiRNAPol2_rhino_KO_46330.bam
ChIPseq_antiRNAPol2_moonshiner_KO_46333.bam
ChIPseq_input_w1118_1_46326.bam
ChIPseq_input_rhino_KO_46329.bam
ChIPseq_input_moonshiner_KO_46332.bam

4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

UCSC browser-compatible bigwig (.bw) files are included in the GEO submission

Methodological details

5. Describe the experimental replicates.

The ChIPseq experiments are supported by complimentary methods:
- Rhino ChIPseq: supported by Rhino IF to assess the typical accumulation in nuclear foci representing germline piRNA clusters (Mohn et al 2014)
- RNA Pol II ChIPseq: supported by evaluation of piRNA cluster transcription by RNAseq and quantitative RNA FISH analyses.

6. Describe the sequencing depth for each experiment.

ChIPseq_antiRhino_w1118_1_22682.bam: 43.52 million reads with 26.01 million uniquely mapped
ChIPseq_antiRhino_Rhino_rhino_KO_22683.bam: 53.33 million

reads with 41.99 million uniquely mapped
ChIPseq_antiRhino_Rhino_moonshiner_KO_22684.bam: 58.43 million reads with 33.99 million uniquely mapped
ChIPseq_antiRhino_Pld_promoter_deletion_43186.bam: 30.7 million reads with 12.09 million uniquely mapped
ChIPseq_antiRNAPol2_w1118_1_46327.bam: 30.85 million reads with 21.52 million uniquely mapped
ChIPseq_antiRNAPol2_rhino_KO_46330.bam: 28.14 million reads with 20.59 million uniquely mapped
ChIPseq_antiRNAPol2_moonshiner_KO_46333.bam: 30.7 million reads with 23.01 million uniquely mapped
ChIPseq_input_w1118_1_46326.bam: 29.26 million reads with 21.97 million uniquely mapped
ChIPseq_input_rhino_KO_46329.bam: 34.8 million reads with 26.47 million uniquely mapped
ChIPseq_input_moonshiner_KO_46332.bam: 31.37 million reads with 24.01 million uniquely mapped

All libraries were sequence single-end 50 bp

7. Describe the antibodies used for the ChIP-seq experiments.

ChIPseq_antiRhino samples: Anti-Rhino polyclonal antibody produced in Rabbit (Mohn et al 2014)
ChIPseq_antiRNAPol2 sample: anti-RNA Polymerase II, 8WG16 (Abcam, ab819)

8. Describe the peak calling parameters.

Peak calling was not utilized in the study

9. Describe the methods used to ensure data quality.

1. qPCR-based enrichment over input (>100 fold enriched at expected loci relative to genomic background)
2. Visual inspection of data in the genome browser to confirm that the ChIPseq signal in wildtype accumulates as expected based on previous literature
3. For Rhino ChIPseq: inclusion of Rhino ChIPseq from Rhino null flies as a control for background IP signal.

10. Describe the software used to collect and analyze the ChIP-seq data.

ChIPseq reads were trimmed to high quality bases 5-45 before mapping to the Drosophila melanogaster genome (dm6, r6.10) using Bowtie (release 0.12.9) with 0-mismatch tolerance. Reads were then computationally extended to 300 nt, reflecting an estimated median DNA fragment length. Normalization between samples was done based on the number of genome-unique mapping reads for each sample. Subsequent quantification of reads mapping to 1 kb tiles was done using bedtools, while relative quantification and plotting was done in R. Rhino ChIP-seq tile signal was normalized to the estimated mappability scores for each 1 kb window, while for Pol II ChIP-seq normalization was done by quantile normalization using the preprocessCore R package. This normalization is under the assumption the Pol II occupancy does not change globally in any of the assayed genotypes (justified by the observed completion ovary development in all genotypes). A pseudo-count of 1 was then added to each tile value before calculation of log2 fold-change values relative to control genotype samples.