

## Supplementary figure legends

### Table S1. Results of BLAST analysis using the predicted polypeptides of the three putative subunits of RNase H2 found in the genome of *Trypanosoma brucei*.

**Figure S1. Protein sequence analysis of putative *T. brucei* type 2 RNase H proteins.** A) Structural models of putative RNase H2 subunits A (Tb427.10.5070), B (Tb427.01.4220) and C (Tb427.01.4730), as predicted by InterPro analysis. Conserved domains are shown, and active site and catalytic residues are highlighted as white boxes and red diamonds above predicted domains, respectively. B) Amino acid alignments of putative RNase H2A catalytic proteins from *T. brucei*, *T. cruzi*, *L. major*, *E. coli* and *T. maritima*. Predicted sequences of the RNase H (H-)domain is highlighted in blue, active site residues in cyan, and catalytic residues are indicated by red dots.

**Figure S2. TbRH2A is a nuclear protein in bloodstream form *T. brucei* parasites.** A) Western blot detection, using anti-HA antiserum, of C-terminally 6HA-tagged TbRH2A protein expressed from its endogenous locus in clonal bloodstream form *T. brucei*. A sample from untagged wild-type (WT) parasites is shown for comparison along with the expected protein size (kDa). B) Representative immunofluorescent images of *T. brucei* parasites expressing TbRH2A-6HA; WT cells are shown for comparison. Anti-HA signal is shown in magenta and DAPI staining in cyan. The cell outline is shown by differential interference contrast microscopy (DIC). Scale bar, 5 $\mu$ m. C) Super-resolution structured illumination imaging of TbRH2A and nuclear DNA (magenta and cyan respectively in the merged, uncoloured images), detected with anti-HA antiserum and DAPI respectively. A respective cell of each cell cycle stage is shown along with graphs plotting length across the nucleus (x-axis, pixels) versus mean pixel intensity at each point (y, arbitrary units) for DAPI (cyan) and TbRH2A (magenta). Scale bars, 5 $\mu$ m. D) Graphs depicting mean fluorescence intensity (arbitrary units, a.u.) of DAPI (left, cyan) and anti-HA- staining (right, magenta) to detect TbRH2A expression. Cells are separated by discernible cell cycle stages (determined by number of nuclear (N) and kDNA (K) signals visualised by DAPI staining; 1N1K, 1N1elongatedK (1N1eK), 1N2K and 2N2K). Each data point denotes the mean fluorescence of an individual cell and the median values (horizontal lines) and interquartile range (error bars) are shown. Significance was determined by the Kruskal-Wallis non-parametric test: (\*\*\*) p-value <0.001, (\*\*\*\*) p-value < 0.0001.

**Figure S3. DNA synthesis occurs in the presence of DNA damage upon depletion of TbRH2A in *T. brucei*.** A) Immunofluorescence detection of nucleic acid (DAPI staining; cyan), thymidine analogue EdU incorporation (via Click-IT chemistry; yellow) and  $\gamma$ H2A (anti- $\gamma$ H2A antiserum; magenta) in TbRH2A RNAi parasites grown in either the absence (tet -) or presence of tet-induction (tet +) for 12, 24 or 36 hr. EdU and  $\gamma$ H2A signal colocalization within several cells is also shown as merged images.

Scale bars, 5 $\mu$ m. B) SR-SIM images of DAPI,  $\gamma$ H2A and EdU staining are shown, along with 3D reconstructions (model), of TbRH2A RNAi cells grown in the absence (tet -) or presence (tet +) of tet-induction. Scale bars, 1  $\mu$ m.

**Figure S4. DNA synthesis continues after depletion of TbRH2A in *T. brucei*.** Graphs show fluorescence intensity of the nucleus after DAPI (left, blue) and EdU- staining (right, orange) at different time points after growth with (Tet+) or without (Tet-) induction of RNAi against TbRH2A. Each dot represents the mean pixel intensity (arbitrary units) of the nucleus of an individual cell; median values (horizontal lines) and interquartile range (error bars) are shown.

**Figure S5. DRIP-seq signal across the 11 Mb-sized chromosomes of *T. brucei* before and after TbRH2A depletion.** DRIP-seq signal for TbRH2A RNAi parasites grown without (blue) and with (orange) tet-induction for 24 hr is plotted as fold-change of IP sample coverage relative to input sample coverage (scale: 1-4 fold change). Upper track shows genes encoded in the sense (black) and antisense (red) directions. Known centromeres are shown as peach ovals. Lowest track shows predicted tandem repeat sequences.

**Figure S6. Distribution of DRIP enriched regions across the 11 Mb-sized chromosomes.** The distribution of DRIP-seq enriched regions (> 1.2 fold change DRIP signal relative to input signal) in TbRH2A RNAi tet-induced (tet +) and uninduced (tet -) parasites. The composition of the Mb-sized chromosomes is shown to the left for comparison.

**Figure S7. Distribution of DRIP enriched regions across the RNA Pol II transcribed polycistronic units.** A) The distribution of DRIP enriched regions between CDS, 5' UTR, 3' UTR and intergenic sequences within the RNA Pol II transcribed PTUs is shown for WT, TbRH2A RNAi uninduced (tet -) and induced (tet +) cells. Total numbers of DRIP enriched regions defined are displayed below plots. B) DRIP-seq signal coverage (normalised to input sample coverage; scale 1-3 fold-change) across an example regions of a PTU within Mb chromosome 2, for TbRH2A RNAi uninduced (tet -) and induced (tet +) samples. CDS regions are indicated by thick black lines below tracks, UTR sequences are indicated by thin black lines, and white arrows indicate transcription direction.

**Figure S8. DRIP-seq signal is enriched in regions flanking CDS of RNA Pol II transcribed genes.** Heatmaps of DRIP-seq signal (normalised to input signal) across each RNA Pol II transcribed protein-coding CDS, plus 1 kb of upstream and downstream flanking region, are shown for WT, TbRH2A RNAi uninduced (tet -) and induced (tet +) cells. Profiles of the average DRIP-seq of all CDS is shown above each heatmap.

**Figure S9. DRIP-seq signal mapping across the translational start sites of RNA Pol II transcribed genes.** A) Average DRIP-seq signal coverage (normalised to input coverage; x axes) is plotted across 1

kb surrounding the ATG start sites (y axes) of the first gene in each RNA Pol II transcribed PTU (left), and all other genes contained with the PTUs (right). Average signal is shown TbRH2A RNAi uninduced (blue; tet -) and induced (orange; tet +) cells. Shaded areas show SEM. B) As in A, but showing the same data for wild type cells (WT, purple) and *T. brucei* RNase H1 null mutants (green, *Tbrh1*<sup>-/-</sup>).

**Figure S10. Levels of DRIP-seq signal are not altered in DNA replication origins SSRs compared with non-origin SSRs.** A) The average DRIP-seq signal coverage (normalised to input sample coverage) is plot across SSRs, plus 1 kb of upstream and downstream flanking regions, which are known origins of replication (ORI) for TbRH2A uninduced (blue; tet -) and induced (orange, tet +) cells. The 5' and 3' boundaries of each SSR were defined as the end and start of flanking transcripts. Shaded areas represent SEM. An example of DRIP-seq signal coverage is shown to the right of the average profiles plot for one ORI SSR. B) As in A, but for SSRs not defined as origins (nonORI). B) As in A, but showing average DRIP-seq coverage for wild type cells (WT, purple) and *T. brucei* RNase H1 null mutants (green, *Tbrh1*<sup>-/-</sup>).

**Figure S11.  $\gamma$ H2A ChIP-seq signal is specifically enriched at sites of transcription initiation after TbRH2A depletion.** Levels of  $\gamma$ H2A ChIP-seq signal in TbRH2A RNAi induced sample coverage is plotted relative to uninduced sample coverage (each first normalised to input samples) for 24 hr (purple, low track) and 36 hr (green, upper track) time points in all 11 Mb-sized chromosomes. Scale: 1-3 fold-change. Other annotations are as in Fig. S5.

**Figure S12. Gene ontology analysis reveals genes associated with antigenic variation are up-regulated after TbRH2A depletion, and those associated with small molecular biosynthesis pathways are down-regulated.** Gene ontology analysis of biological process terms associated with genes found to be significantly up-regulated or down-regulated in RNA-seq analysis comparing RNA abundance after 36 hr TbRH2A of depletion via RNAi compared with uninduced cells. Parent terms are also depicted. Terms found to be enriched in the upregulated and downregulated genes are coloured green and orange, respectively. Colour intensity indicates significance, which was determined as adjusted p value. Terms with values > 0.01 were deemed not significant (white).

**Figure S13. DRIP-seq signal is enriched across the BESs after TbRH2A depletion.** DRIP-seq signal coverage (normalised to input coverage) is plotted across the BESs for WT (pink), TbRH2A uninduced (blue; tet -) and induced (orange, tet +) cells. The lowest track shows the structure of each BES; promoters (cyan), ESAGs (blue), 70-bp repeats (purple), VSGs (red) and other genes (green) are annotated as boxes. Black circles denote the end of the BES sequence assembly.

**Figure S14.  $\gamma$ H2A ChIP-seq is enriched across the BESs after TbRH2A depletion.** Levels of  $\gamma$ H2A ChIP-seq signal in TbRH2A RNAi induced sample coverage is plotted relative to uninduced sample

coverage (each first normalised to input samples) for 24 hr (purple, lower track) and 36 hr (green, upper track) time points, across each BESs. Scale, 1-3 fold-change. The lowest track shows the structure of each BES; promoters (cyan), ESAGs (blue), 70-bp repeats (purple), VSGs (red) and other genes (green) are annotated as boxes. Black circles denote the end of the BES sequence assembly.

**Figure S15. DRIP-seq and  $\gamma$ H2A CHIP-seq signal is enriched across rRNA genes after TbRH2A depletion.** Upper diagrams shows DRIP-seq signal coverage (normalised to input coverage) plotted across the rRNA locus on chromosome 3 for TbRH2A uninduced (blue; tet -) and RNAi induced (orange, tet +) cells. CDS are indicated as black boxes. The lower diagrams shows the same rRNA locus but depicts levels of  $\gamma$ H2A CHIP-seq signal, plotting coverage in TbRH2A RNAi induced samples relative to coverage in uninduced samples (each first normalised to input samples) for 24 hr (purple) and 36 hr (green) time points (scale, 1-3 fold-change)

**Figure S16. Silent BES-housed VSGs are transcribed after TbRH2A depletion.** RNA-seq read coverage is plotted over the coding regions of VSGs housed within silent BESs for two independent replicates (blue and orange) of TbRH2A RNAi parasites grown for 36 hr in the absence (tet -) or presence (tet +) of tet-induction. The VSG names and BESs in which they are contained are indicated.

Query ID	Hit Species	Hit Annotation	Accession number	E-value	Identity score
Tb427.10.5070	<i>Trypanosoma congolense</i>	Putative RNase H	CCC93661.1	0	539
	<i>Trypanosoma cruzi</i>	RNase H	XP_806655.1	4.00E-177	502
	<i>Leishmania major</i>	Putative RNase H	XP_001686636.1	1.00E-116	348
	<i>Leishmania donovani</i>	RNase H, putative	XP_003865186.1	2.00E-115	339
	<i>Mus musculus</i>	RNase H2 A subunit	NP_081463.1	3.00E-47	164
	<i>Arabidopsis thaliana</i>	Polynucleotidy transferase, ribonuclease-H like superfamily	NP_565584.1	3.00E-46	161
	<i>Homo sapiens</i>	RNase H2 A subunit	NP_006388.2	4.00E-45	159
	<i>Drosophila melanogaster</i>	Uncharacterised protein	NP_608521.1	7.00E-40	146
	<i>Plasmodium falciparum</i>	RNase H2 A subunit, putative	XP_966221.1	3.00E-40	145
	<i>Saccharomyces cerevisiae</i>	RNase H2 catalytic subunit RNH201	NP_014327.1	6.00E-40	145
	<i>Caenorhabditis elegans</i>	RNase H2 subunit A	NP_495796.1	4.00E-39	143
	<i>Escherichia coli</i>	RNase HII	NP_414725.1	1.00E-04	46.2
Tb427.01.4220	<i>Trypanosoma congolense</i>	conserved hypothetical protein	CCC89404.1	8.00E-143	421
	<i>Trypanosoma cruzi</i>	hypothetical protein	XP_808292.1	5.00E-103	322
	<i>Danio rerio</i>	ribonuclease H2 subunit B isoform X1	XP_021334305.1	7.00E-13	72.4
	<i>Mus musculus</i>	ribonuclease H2 subunit B	NP_080277.1	8.00E-04	45.1
	<i>Homo sapiens</i>	ribonuclease H2 subunit B isoform X3	XP_011533532.1	0.33	36.6
Tb427.01.4730	<i>Trypanosoma brucei equiperdum</i>	Ribonuclease H2 non-catalytic subunit	RHW74391.1	8.00E-113	326
	<i>Trypanosoma congolense</i>	unnamed protein product	CCD13112.1	1.00E-50	168
	<i>Trypanosoma cruzi</i>	hypothetical protein	XP_809724.1	2.00E-43	149
	<i>Leishmania braziliensis</i>	Ribonuclease_H2_non-catalytic_subunit	SYZ63788.1	7.00E-19	87.8
	<i>Daphnia magna</i>	Ribonuclease H2 subunit C	KZS21140.1	0.27	41.2

Table. S1

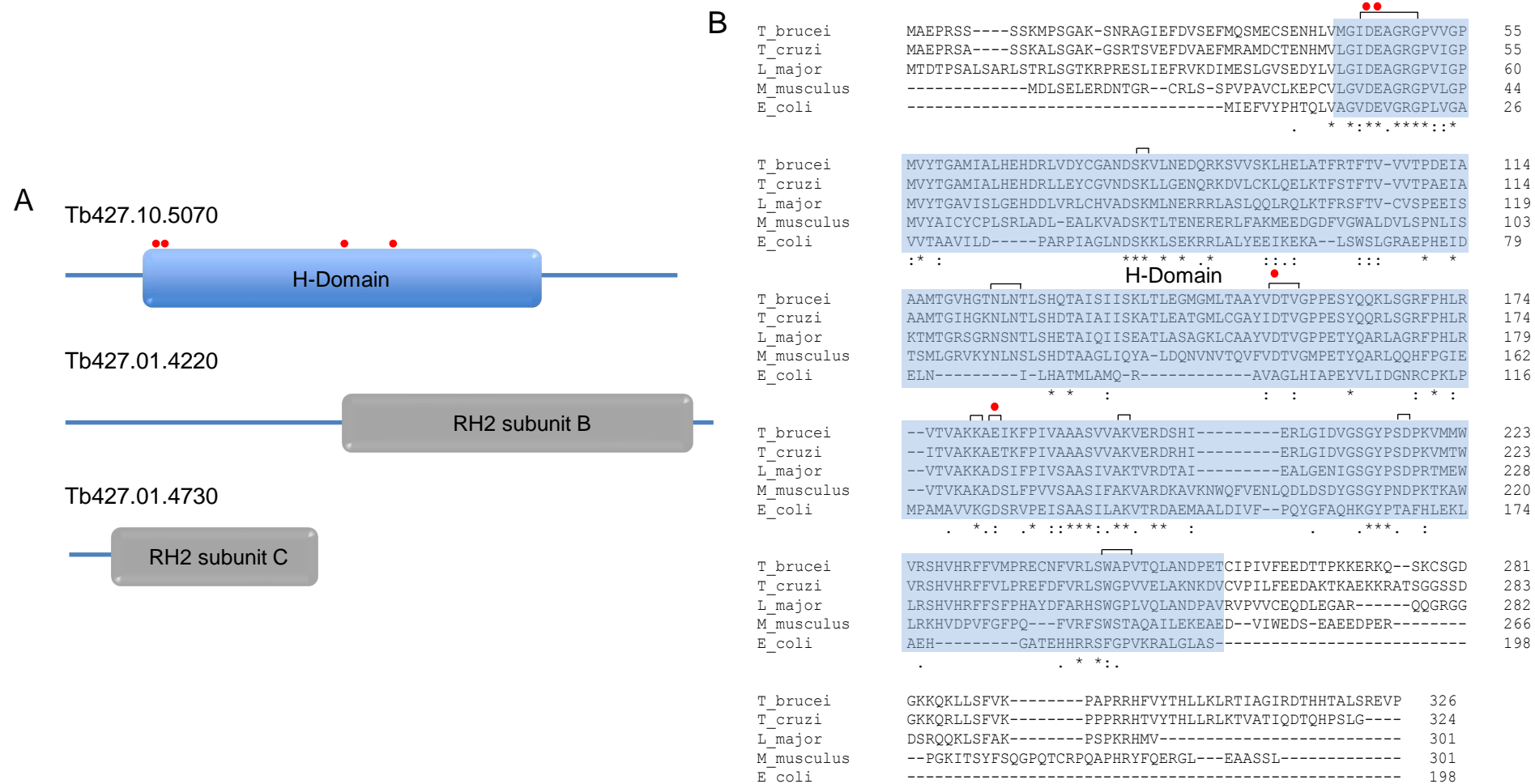


Fig. S1

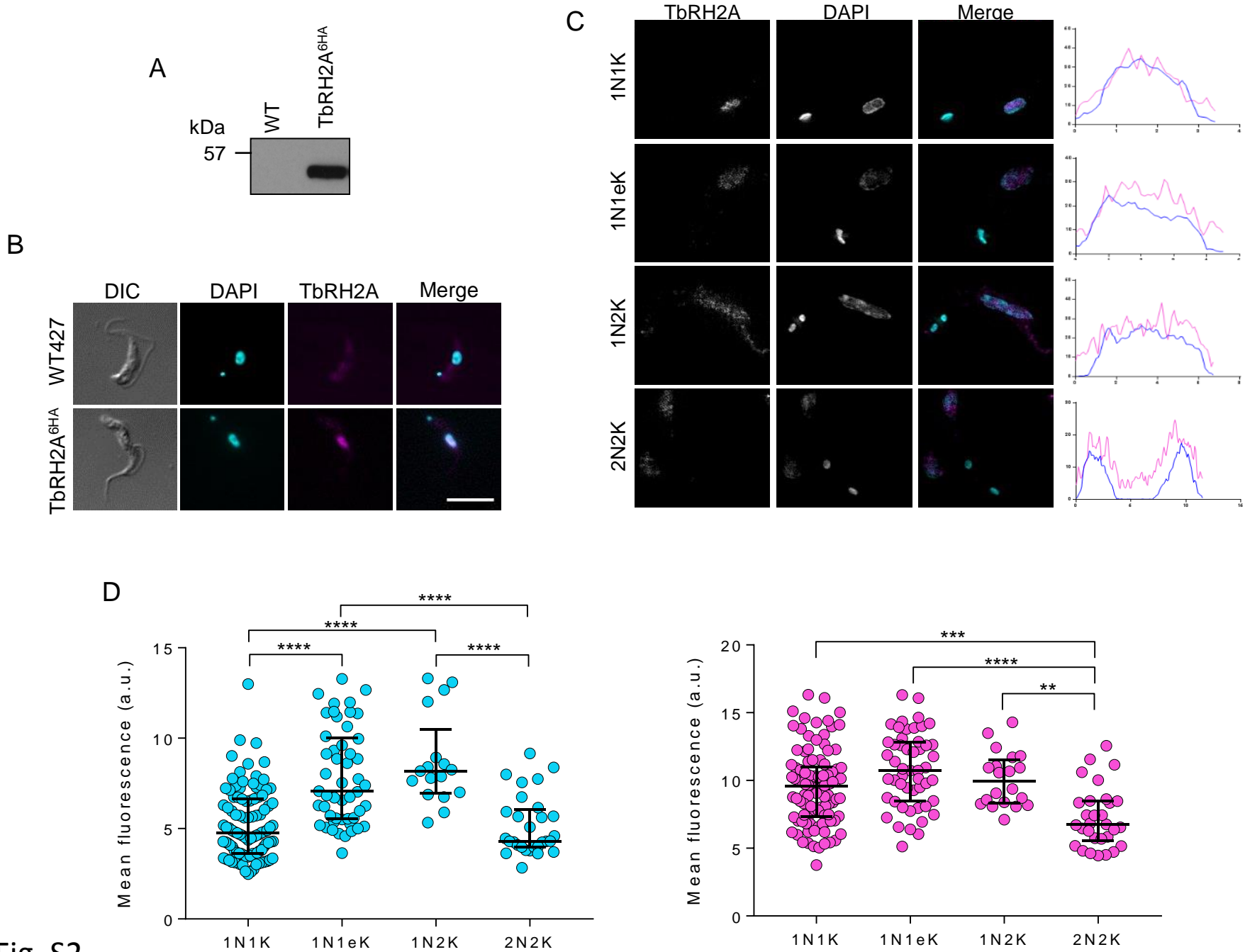
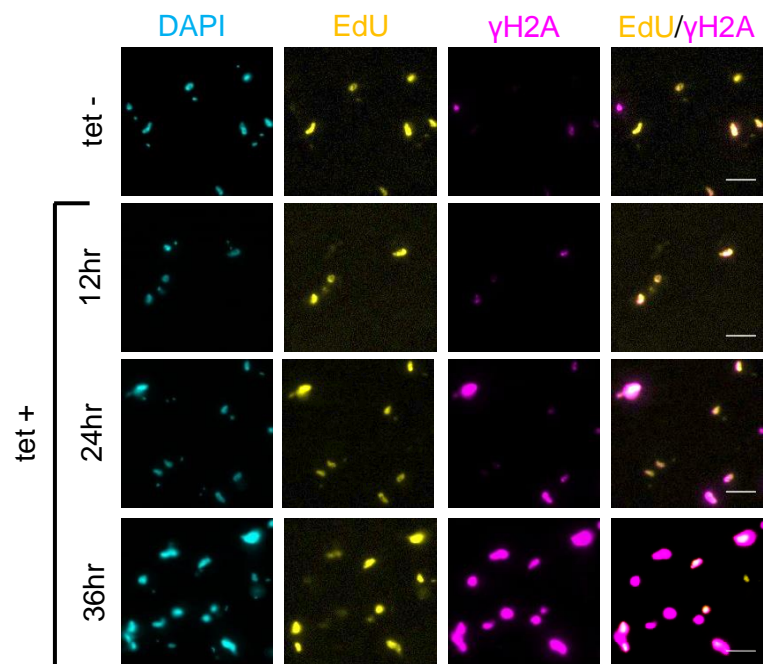


Fig. S2

A



B

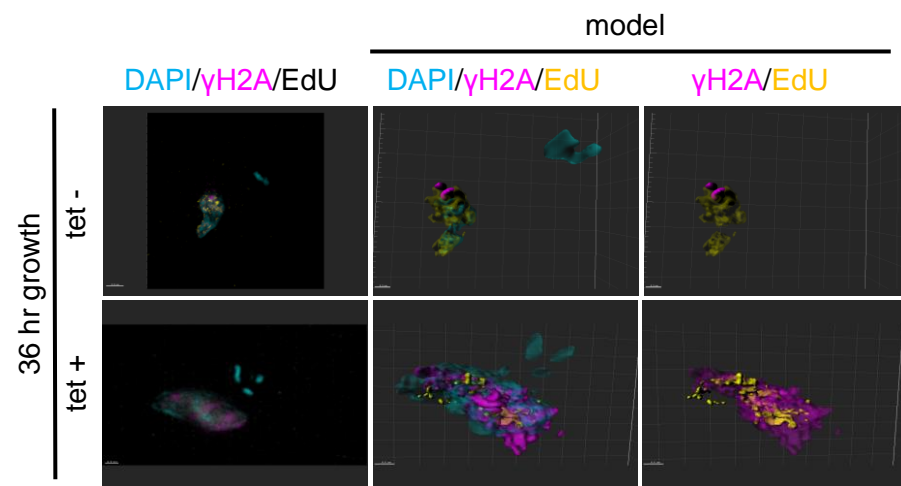


Fig. S3



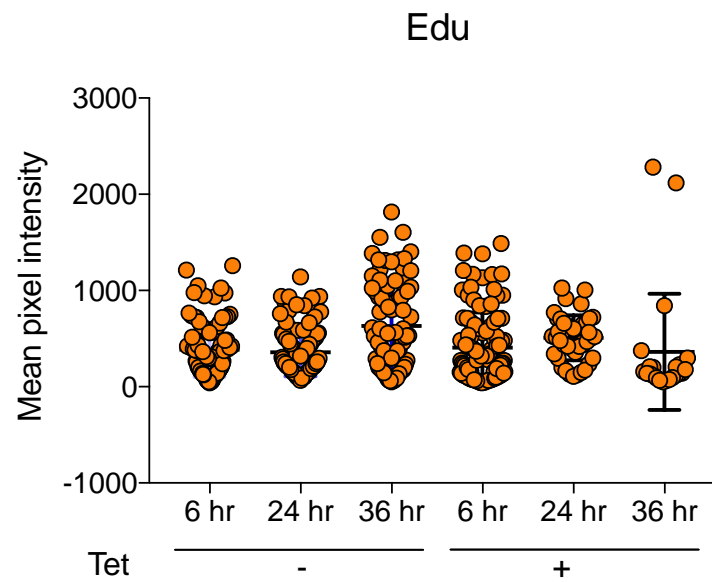
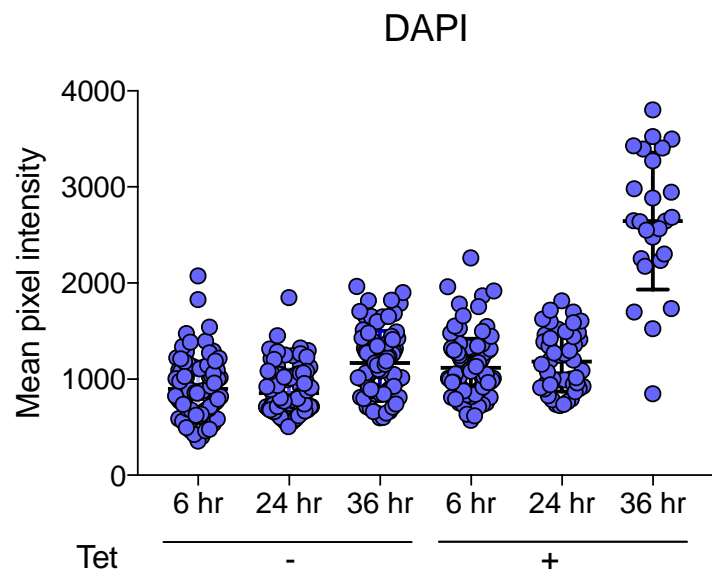


Fig. S4

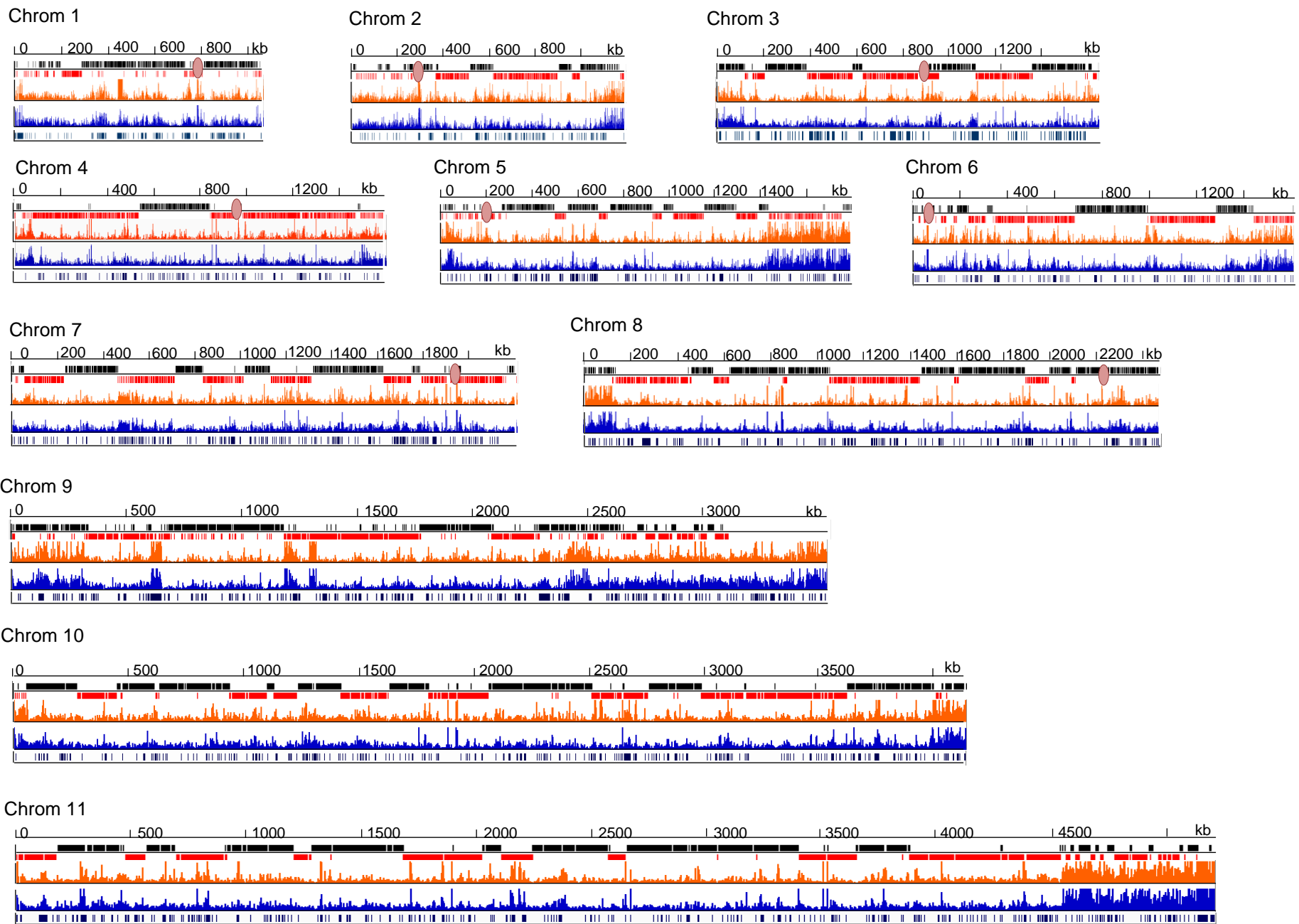
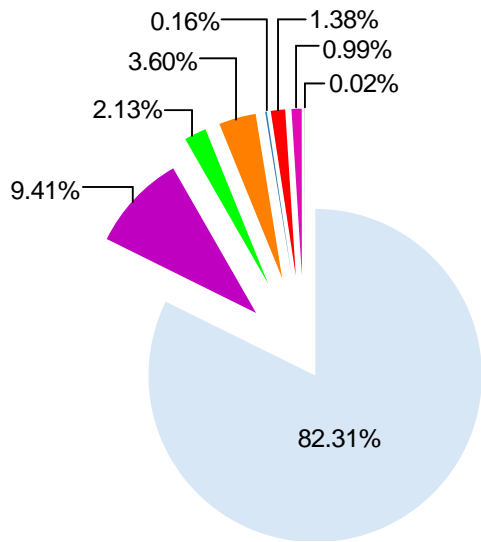
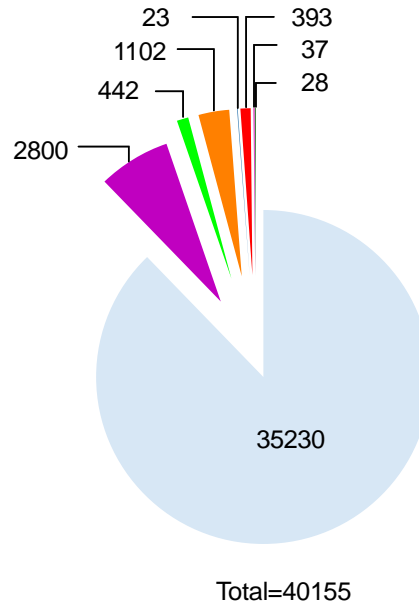


Fig. S5

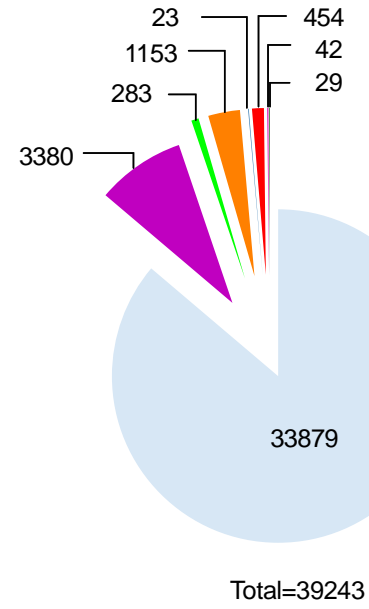
Mb Chromosome composition



tet -



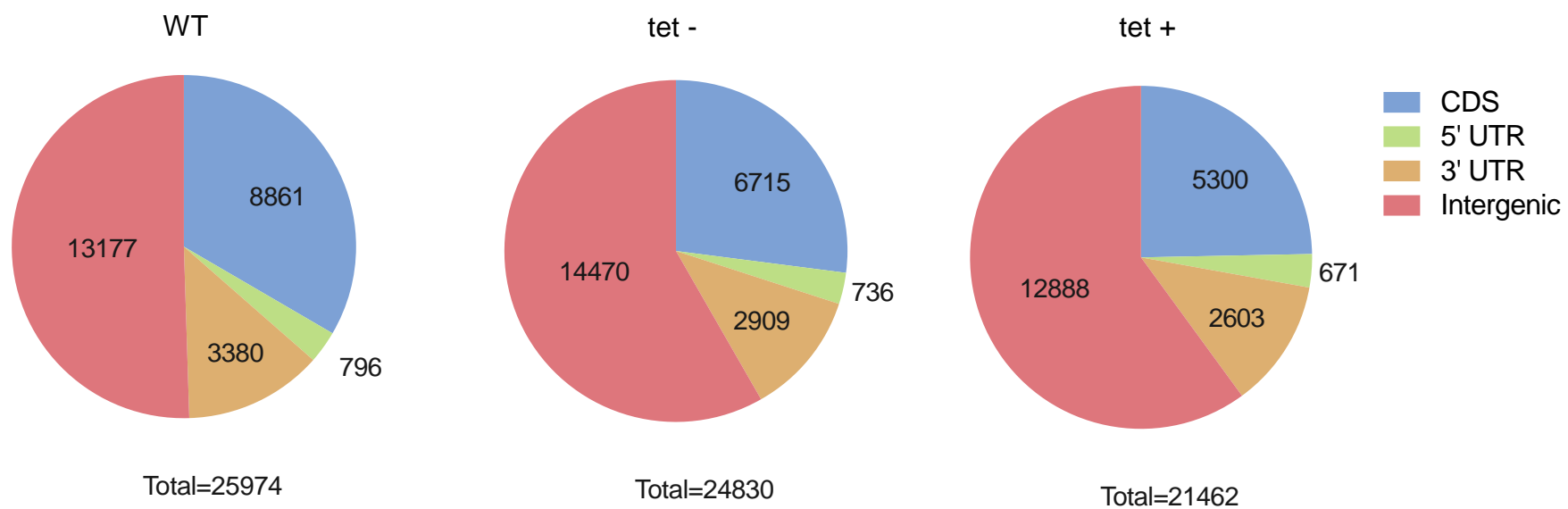
tet +



- Pol II PTU
- VSG array
- Strandswitch regions
- Pol I genes
- Pol III genes
- Retrotransposons
- Centromeres
- SL RNA

Fig. S6

A



B

Tb927\_02\_v5.1: 395,532 - 423,860

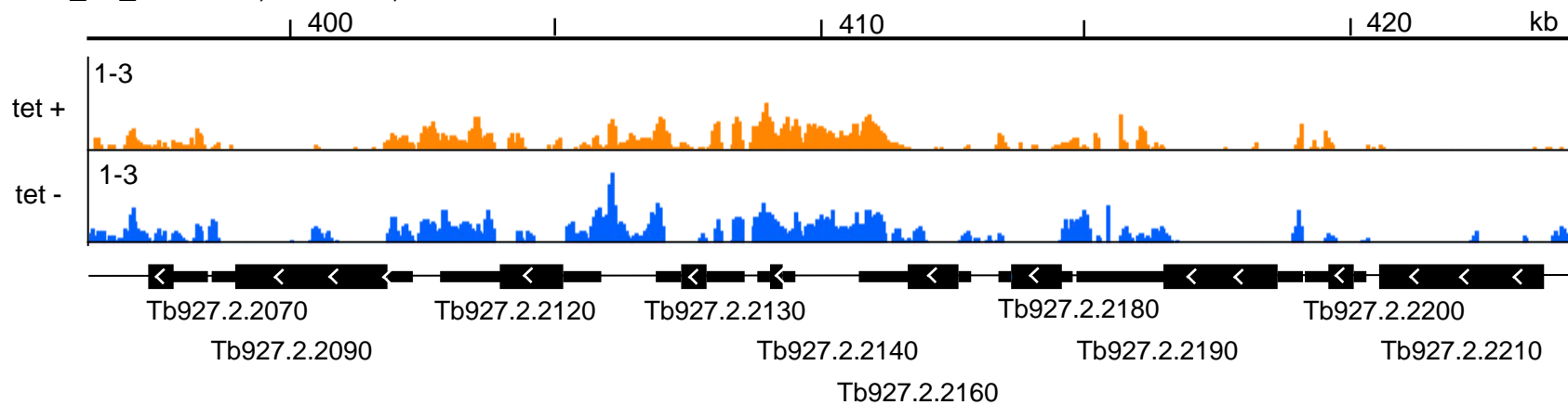


Fig. S7

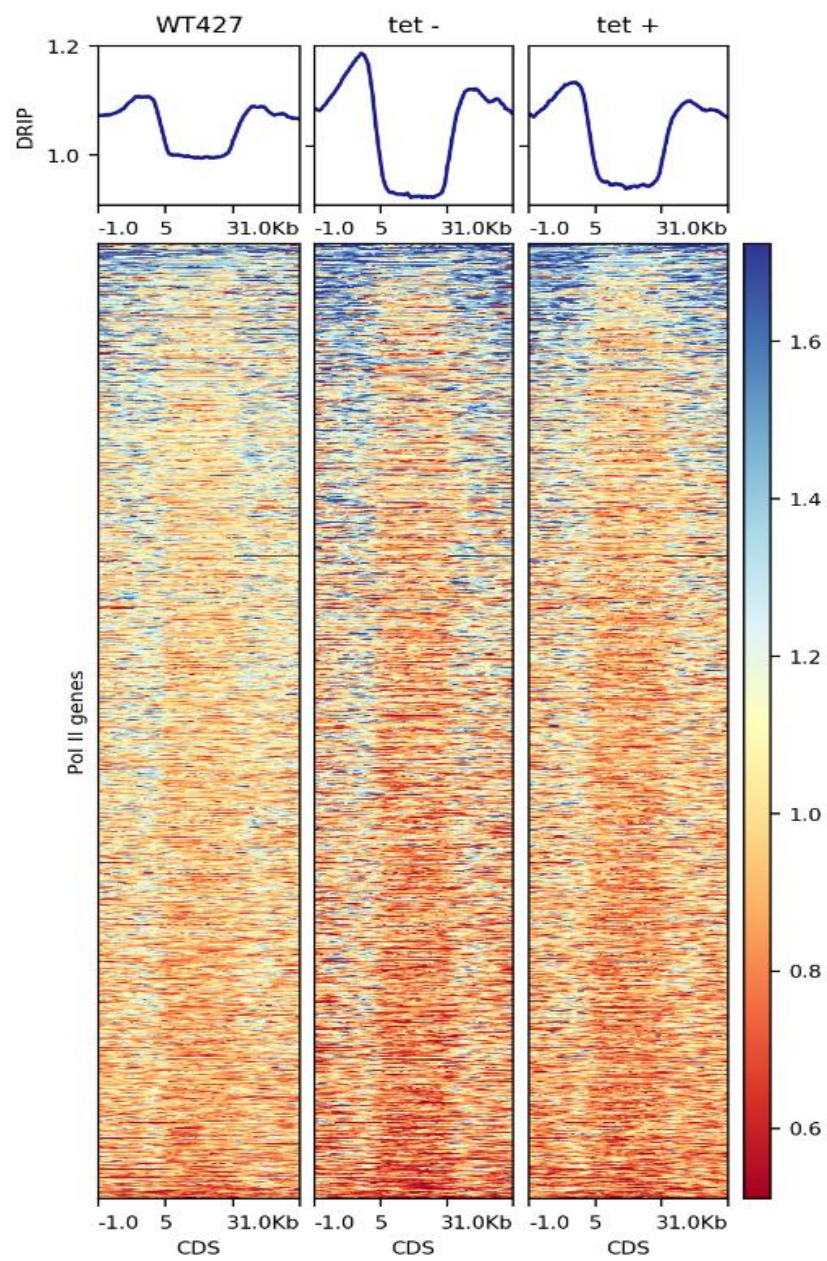


Fig. S8

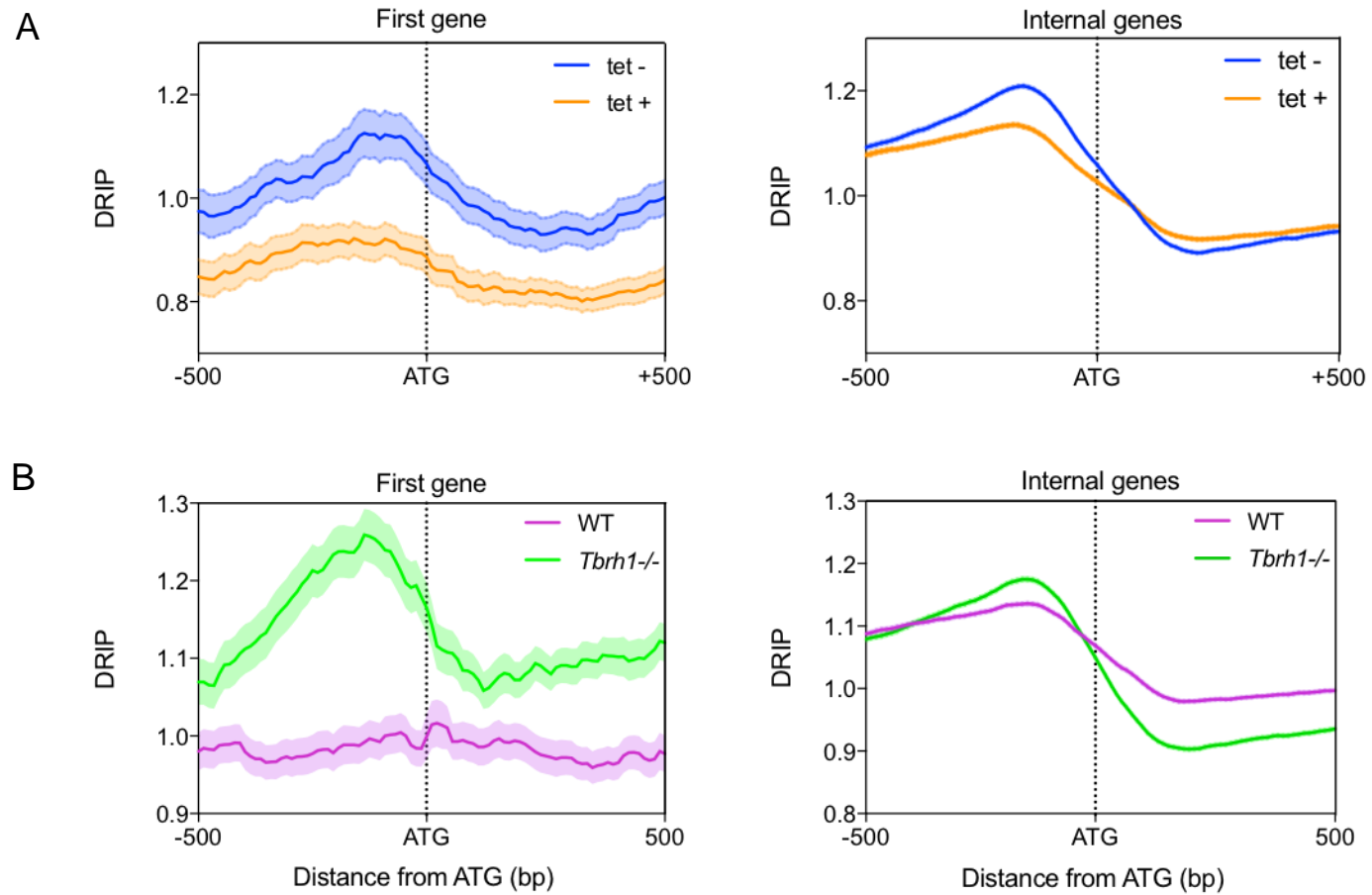


Fig. S9

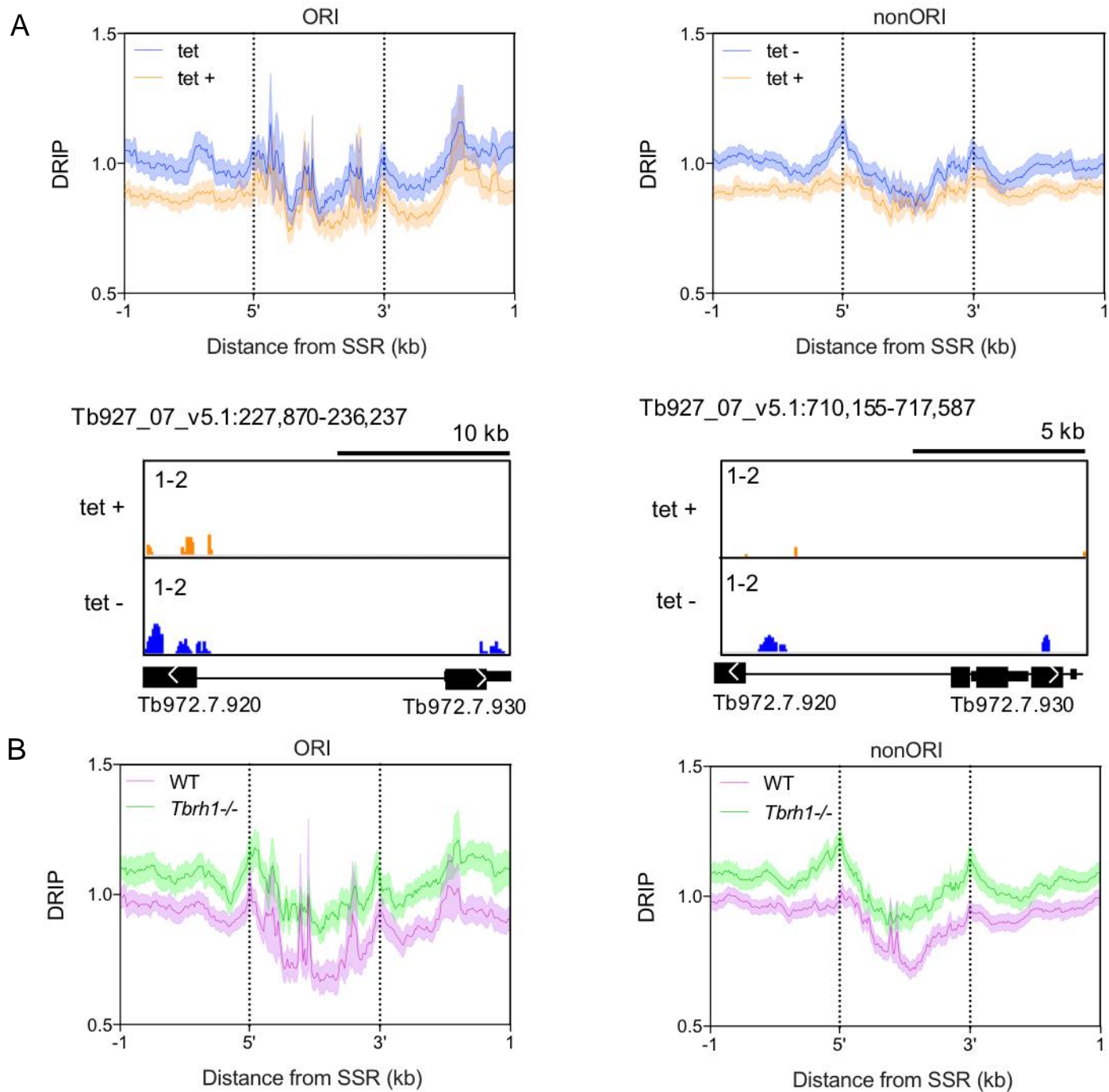


Fig. S10

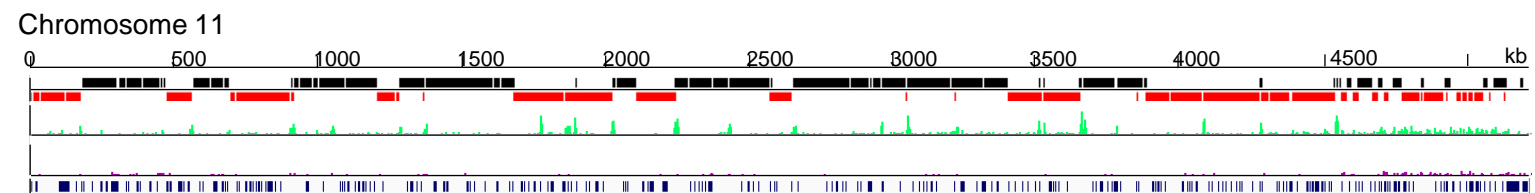
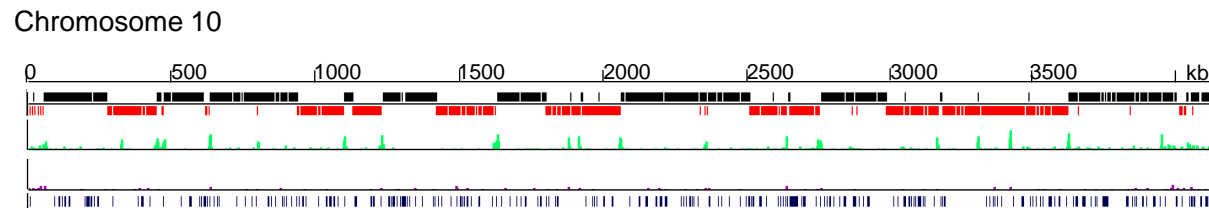
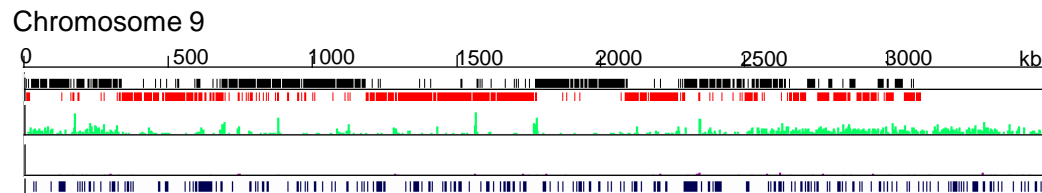
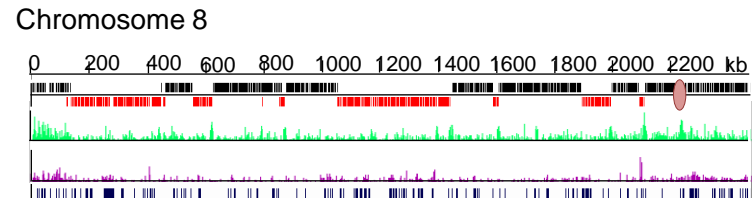
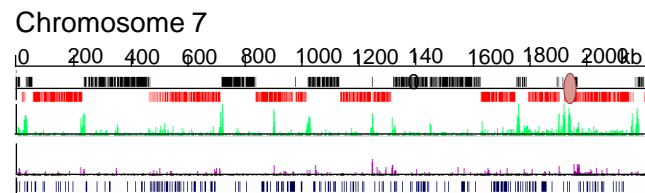
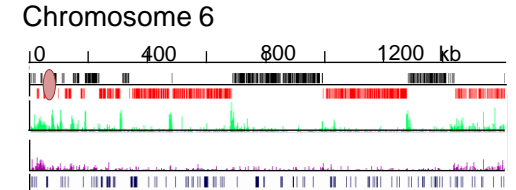
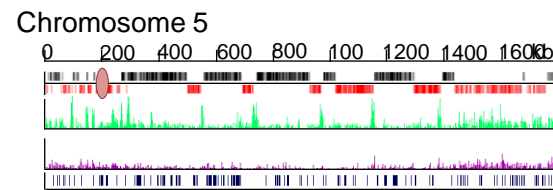
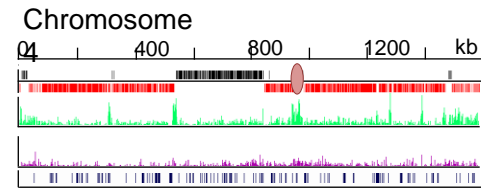
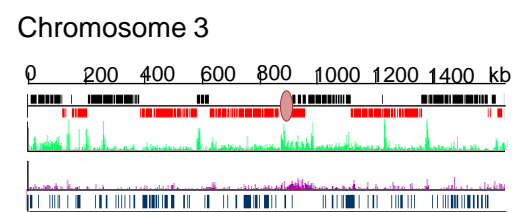
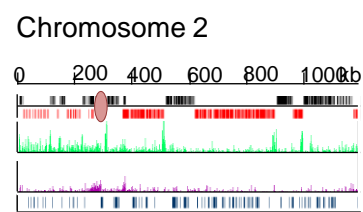
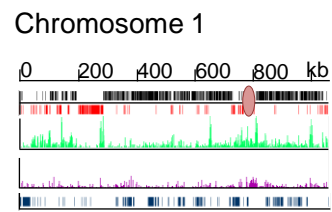


Fig. S11



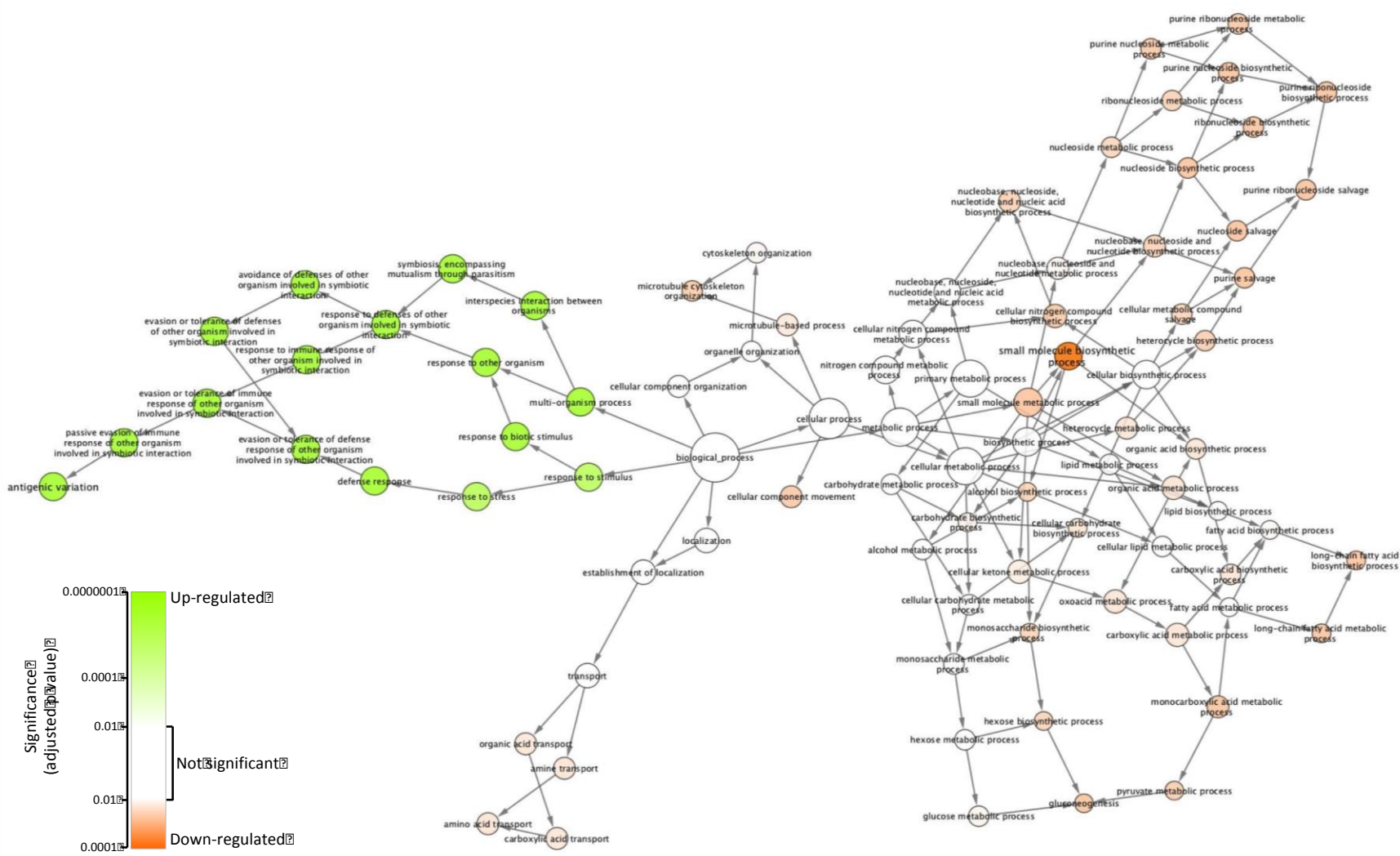


Fig. S12

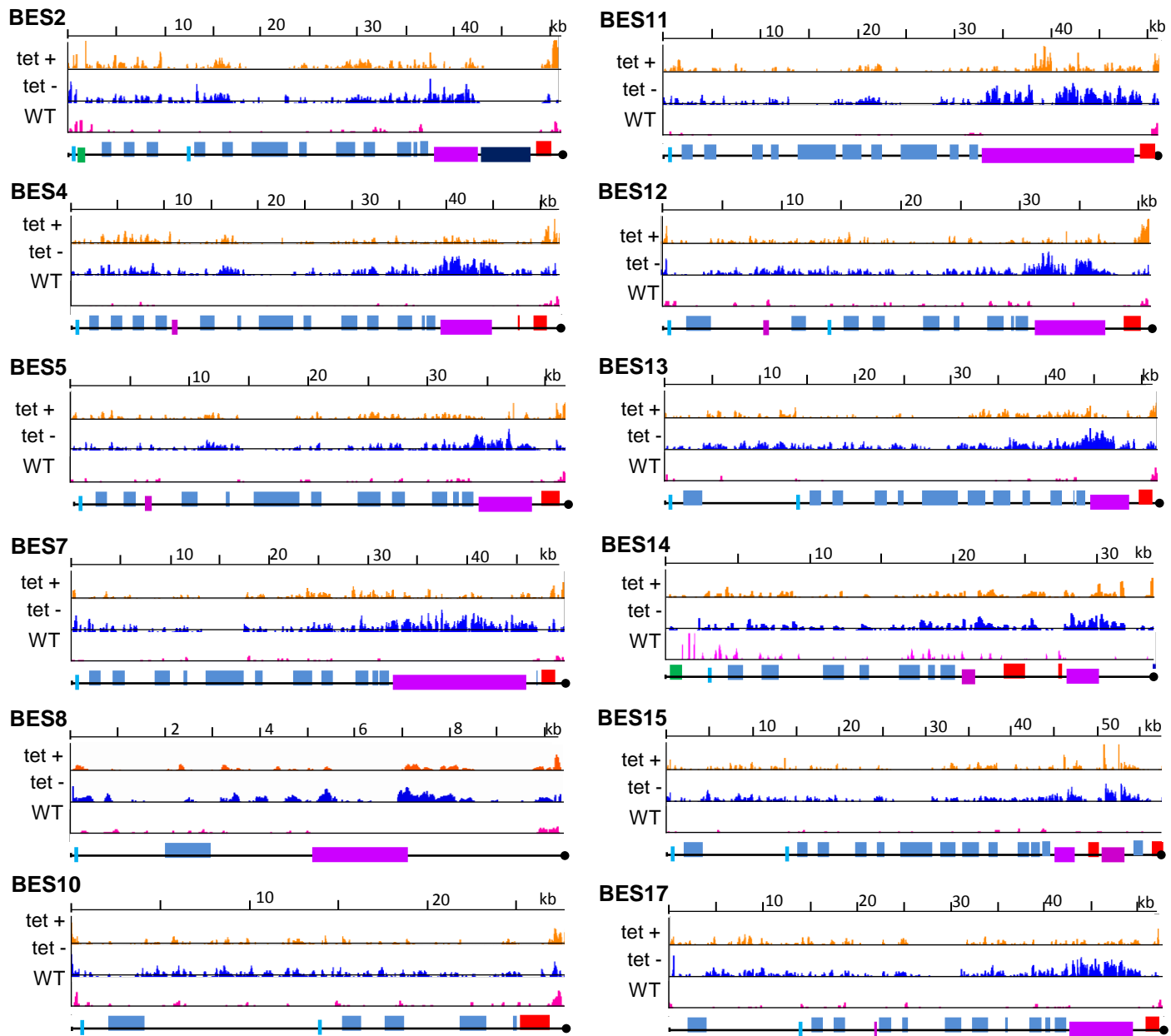


Fig. S13

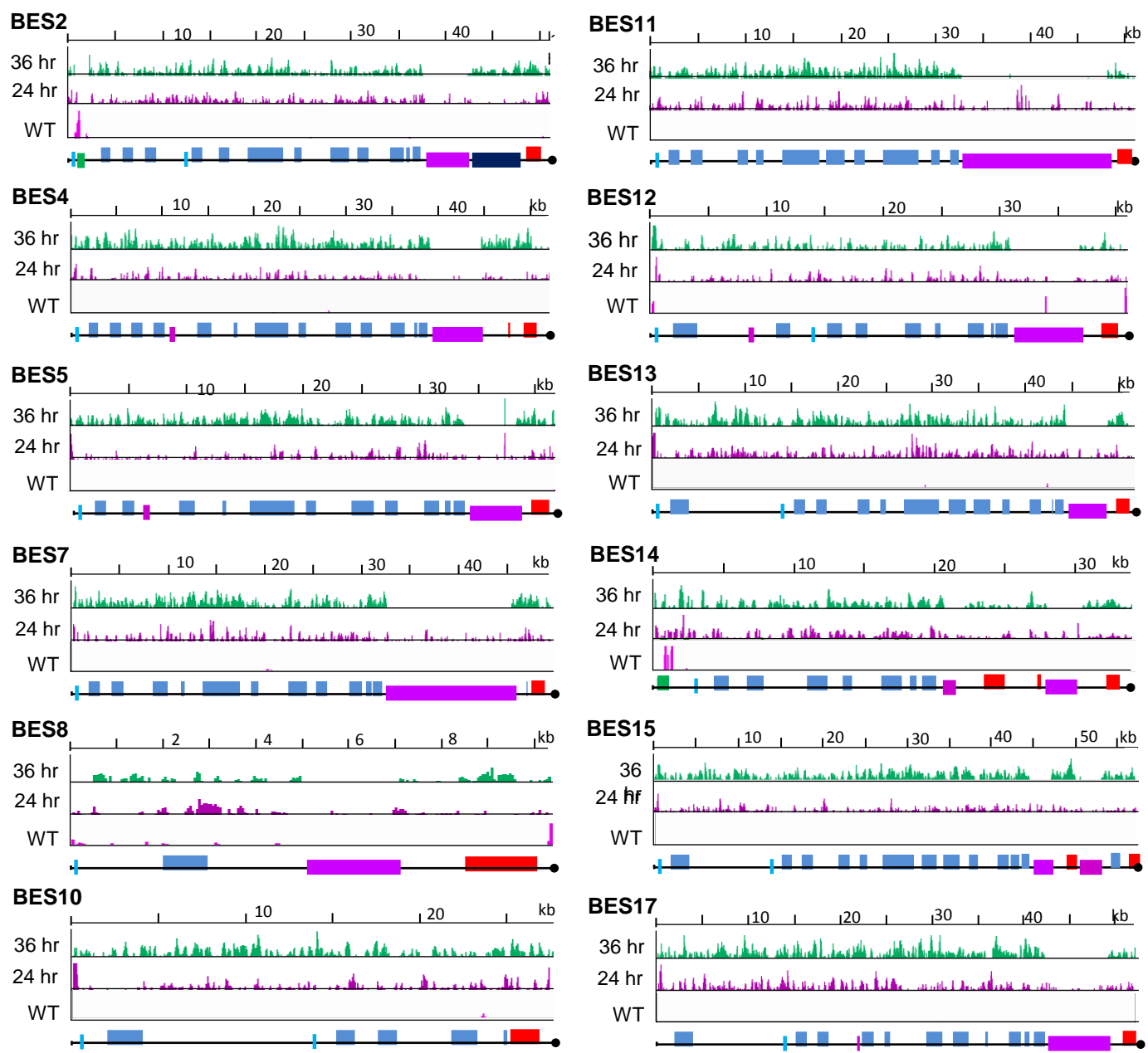


Fig. S14

Tb927\_03\_v5.1:899,998-969,489

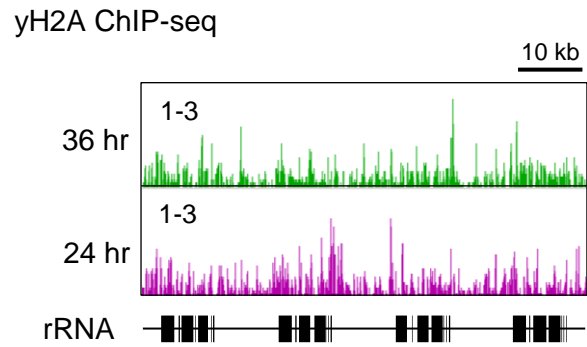
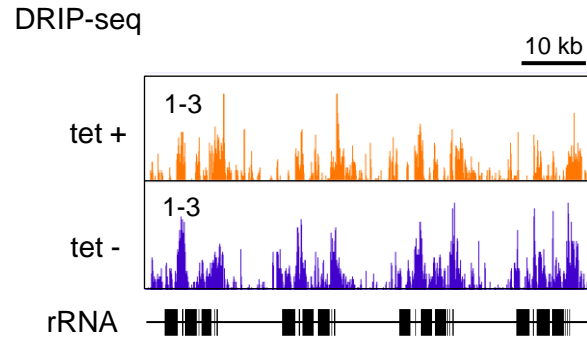
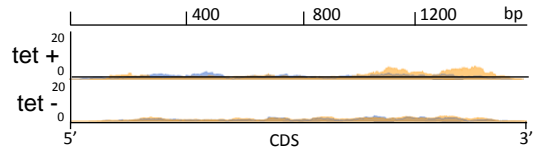
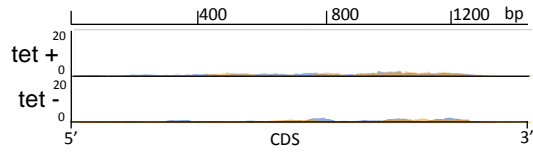


Fig. S15

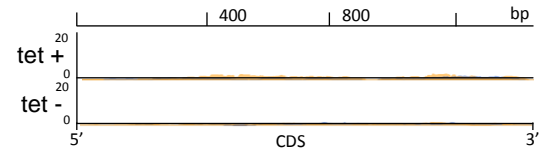
**VSG9 (BES2)**



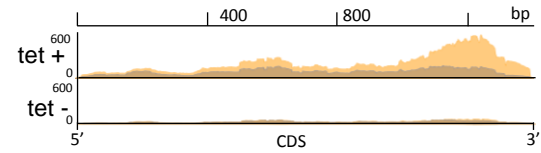
**VSGT3 (BES4)**



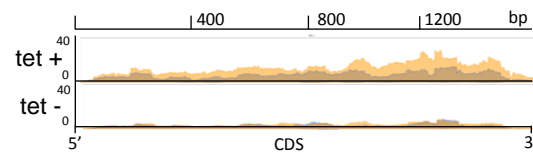
**VSG16 (BES11)**



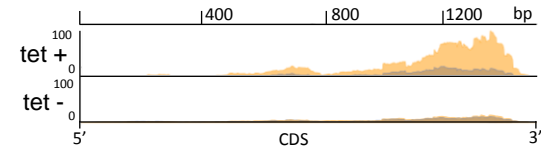
**VSG8 (BES12)**



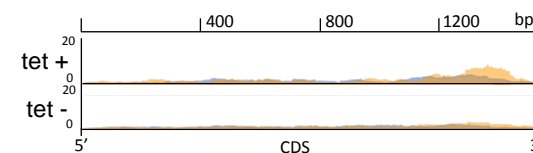
**VSG800 (BES5)**



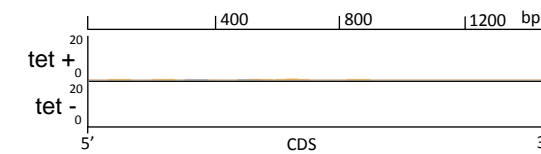
**VSG17 (BES13)**



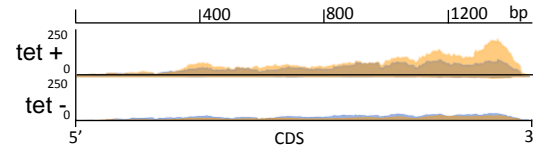
**VSG224 (BES7)**



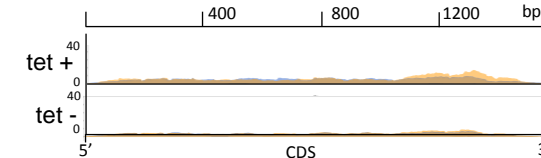
**VSG19 (BES14)**



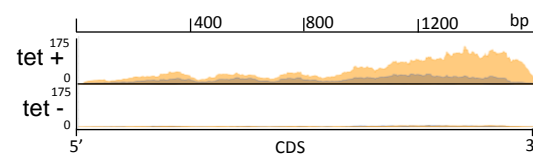
**VSG14 (BES8)**



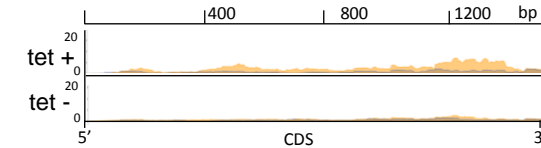
**VSG11 (BES15)**



**VSG15 (BES10)**



**VSG13 (BES17)**



**Fig. S16**