Antigenic variation by switching inter-chromosomal interactions with an RNA splicing locus in trypanosomes

Authors: Joana Faria[1¶], Vanessa Luzak[2,3¶], Laura S.M. Müller[2,3¶], Benedikt G. Brink[2,3], Sebastian Hutchinson[1†], Lucy Glover[1#], David Horn[1*], T. Nicolai Siegel[2,3*]

Affiliations:

[1]The Wellcome Trust Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK.

[2]Department of Veterinary Sciences, Experimental Parasitology, Ludwig-Maximilians-Universität München, Lena-Christ-Str. 48, 82152 Planegg-Martinsried, Germany

[3]Biomedical Center Munich, Department of Physiological Chemistry, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Planegg-Martinsried, Germany

[†]Current address: Trypanosome Cell Biology Unit, INSERM U1201, Department of Parasites and Insect Vectors, Institut Pasteur, 25-28 Rue du Docteur Roux 75015, Paris, France.

[#]Current address: Trypanosome Molecular Biology, Department of Parasites and Insect Vectors, Institut Pasteur, 25-28 Rue du Docteur Roux 75015, Paris, France.

¶ These authors contributed equally

* Correspondence: NS, n.siegel@lmu.de; DH, d.horn@dundee.ac.uk

Keywords: 3D genome architecture; RNA maturation; antigenic variation; monoallelic; *Trypanosoma brucei*.

Abstract

Highly selective gene expression is a key requirement for antigenic variation in several pathogens, allowing evasion of host immune responses and maintenance of persistent infections. African trypanosomes, parasites that cause lethal diseases in humans and livestock, employ an antigenic variation mechanism that involves monogenic antigen expression from a pool of >2500 antigen coding genes. In other eukaryotes, the expression of individual genes can be enhanced by mechanisms involving the juxtaposition of otherwise distal chromosomal loci in the three-dimensional nuclear space. However, trypanosomes lack classical enhancer sequences or regulated transcription initiation and the monogenic expression mechanism has remained enigmatic. Here, we show that the single expressed antigen coding gene displays a specific inter-chromosomal interaction with a major mRNA splicing locus. Chromosome conformation capture (Hi-C), revealed a dynamic reconfiguration of this inter-chromosomal interaction upon activation of another antigen. Super-resolution microscopy showed the interaction to be heritable and splicing dependent. We find that the two genomic loci are connected by the antigen exclusion complex, whereby VEX1 associated with the splicing locus and VEX2 with the antigen coding locus. Following VEX2 depletion, loss of monogenic antigen expression was accompanied by increased interactions between previously silent antigen genes and the splicing locus. Our results reveal a novel mechanism to ensure monogenic expression, requiring the spatial integration of antigen transcription and mRNA splicing in a dedicated compartment. These findings suggest a new means of post-transcriptional gene regulation.

Main Text:

Monogenic expression, the expression of a single gene from a large gene family, is essential for several important biological processes. One of the most striking examples of such regulation is the expression of a single odorant receptor from more than 1400 genes in mammalian olfactory sensory neurons [1]. Likewise, monogenic expression is a key feature of antigenic variation, an immune evasion strategy used by pathogens such as *Plasmodium falciparum* or *Trypanosoma brucei*. Antigenic variation refers to the capacity of an infecting organism to systematically alter the identity of proteins displayed to the host immune system [2]. How pathogens ensure the exclusive expression of only one antigen from a large pool of antigen coding genes remains one of the most intriguing questions in infection biology.

In *T. brucei*, a unicellular parasite responsible for lethal and debilitating diseases in humans and animals, 10 million copies of a single variant surface glycoprotein (VSG) isoform are exposed on the surface of the parasite. The exclusive expression of only one VSG gene per cell and the periodic switching of the expressed VSG gene allow the parasite to evade the host immune system and to maintain persistent infections [3,4]. While the *T. brucei* genome encodes for >2600 VSG isoforms, in the bloodstream of the mammalian host, a VSG gene can only be transcribed when located in one of ~15 VSG expression sites. Those bloodstream expression sites are polycistronic transcription units located adjacent to telomeres on different chromosomes. Each bloodstream expression site contains an RNA polymerase I (Pol I) promoter, followed by several expression site associated genes (ESAGs) and a single VSG gene [5].

Notably, Pol I transcription initiates at all VSG expression site promoters, but transcription elongation and transcript processing are highly selective and limited to just one expression site at a time [6,7]. As a result, the single active VSG gene is expressed as the most abundant mRNA and protein in the cell; 5-10% of the total in each case. Why transcription is aborted at all but one expression site is not known, but it has been shown that only transcripts

from the actively transcribed expression site are processed into mature mRNA. In trypanosomes, mRNA maturation involves *trans*-splicing, a process that adds a common spliced leader sequence to each pre-mRNA and is coupled to polyadenylation [8]. In addition, the proximity of individual genes to nuclear condensates composed of splicing factors has recently been proposed to play a role in gene expression regulation in mammals [9,10]. Thus, regulated access to RNA maturation compartments may represent an evolutionary conserved strategy for gene expression control.

One mechanism to ensure monogenic expression is the juxtaposition of otherwise distal chromosomal loci in the three-dimensional nuclear space. In particular, specific interactions between promoter and enhancer sequence elements can ensure the selective regulation of individual genes. Although classic enhancer structures appear to be absent in many unicellular eukaryotes such as trypanosomes, several observations suggest that a specific genome organization is required for monogenic VSG expression. The single active VSG gene is transcribed in an extranucleolar Pol I compartment known as the expression site body [11]. In those very rare cases ($<10^{-8}$) where two VSG genes are simultaneously active, both co-localize at the expression site body [12,13]. In addition, the transcribed chromosome core regions and the sub-telomeric regions coding for the large reservoir of silent VSG genes, appear to fold into structurally distinct compartments [5], similar to active A and silent B compartments described in mammalian cells [14]. While the nature of the expression site body has remained enigmatic, a protein complex specifically associated with the active VSG gene was identified recently. VSG-exclusion 1 (VEX1) emerged from a genetic screen for allelic exclusion regulators [15] while VEX2 was affinity-purified in association with VEX1 [16]. The bipartite VEX protein complex maintains mutually exclusive VSG expression [16] but it remains unclear how these proteins exert their function. In this study we aimed to identify the mechanism that connects RNA maturation, genome architecture and the VEX complex to ensure monogenic antigen expression.

Given the well-characterized role of promoter-enhancer interactions in the selective regulation of genes, we set out to identify specific DNA-DNA interactions with a regulatory role in monogenic VSG expression. To this end we used a *T. brucei* culture homogenously expressing a single VSG gene for chromosome conformation capture (Hi-C) analysis. In addition, we employed the mHi-C analysis pipeline, which allowed us to retain many multi-mapping reads and greatly increased the read coverage across repetitive regions of the genome [17].

In order to visualize specific interaction patterns of loci of interest (viewpoints) in the Hi-C dataset, we applied a virtual 4C analysis pipeline to extract genome-wide interaction profiles for chosen viewpoints. To identify VSG gene specific interaction patterns, we chose the active and several inactive VSG genes located in expression sites as viewpoints and plotted the extracted virtual 4C interaction data onto the genome. As expected, we observed a distance-dependent decay of intra-chromosomal interactions between each viewpoint and its upstream and downstream genomic region (**Fig. 1a and Extended Data Fig. 1a**).

Strikingly, we found the active *VSG-2* gene located on chr. 6 in expression site 1 to very frequently interact with a single, distinct locus on chr. 9 (**Fig. 1a-b**). Levels of interaction frequency were higher than intra-chromosomal interactions of *VSG-2* with its genomic location on chr. 6, pointing to a strong and stable inter-chromosomal interaction. The locus on chr. 9 interacting with the active VSG gene is the SL-RNA array, a genomic locus essential for RNA maturation. This locus contains a cluster of ~150–200 tandemly repeated genes encoding the spliced leader RNA (SL-RNA). SL-RNA is an RNA Pol II-transcribed ncRNA that is *trans*-spliced to the 5´-end of all trypanosome mRNAs, conferring the 5´-cap structure required for RNA maturation, export and translation [8]. Conversely, VSG genes residing in inactive expression sites interacted less frequently or at background levels with the SL-RNA locus (**Fig. 1a-b and Extended Data Fig. 1a**). In agreement with these observations, when we chose the SL-RNA locus as viewpoint, we found it to interact more frequently with the active

VSG expression site than with any inactive VSG expression site (**Extended Data Fig. 1b**). Thus, the Hi-C analysis revealed a strong and selective interaction between the Pol I-transcribed active VSG gene and the Pol II-transcribed SL-RNA locus located on a different chromosome.

To visualize the spatial proximity between the active VSG gene and the SL-RNA locus at the level of individual cells and with an independent assay, we performed super-resolution immunofluorescence microscopy. The site of VSG transcription is characterized by an extra-nucleolar accumulation of RNA Pol I [11]. The site of Pol II-transcribed SL-RNA is marked by an accumulation of the small nuclear RNA-activating protein complex (tSNAPc), an RNA Pol II promoter-binding transcription factor [18]. Both the VSG transcription and the SL-RNA transcription compartments appear to have diameters of approx. 300 nm, within a *T. brucei* nucleus with a diameter of approx. 2 µm. Super-resolution microscopy revealed one VSG transcription compartment, reflecting the hemizygous active sub-telomeric *VSG-2*, and two separate SL-RNA transcription compartments, reflecting the diploid SL-RNA arrays located in the core of chr. 9 (**Fig. 1c**). By scoring nuclei for overlapping, adjacent or separate VSG and SL-RNA transcription compartments, we found that one of the SL-RNA transcription compartments was adjacent to the VSG transcription compartment in the majority of cells (**Fig. 1c**). However, during DNA replication in S phase, the VSG and SL-RNA transcription compartments were detected in separate locations in >50% of nuclei (**Fig. 1c and Extended Data Fig. 1c**). Therefore, throughout this study, immunofluorescence assays (IFAs) were subsequently performed in G1 cells, unless indicated otherwise. Taken together, IFAs supported the findings made by Hi-C, suggesting that the Pol I VSG transcription compartment interacts with one of the Pol II SL-RNA transcription compartments. Further, they suggest that the interaction between both compartments is resolved during S phase and successfully re-established after replication.

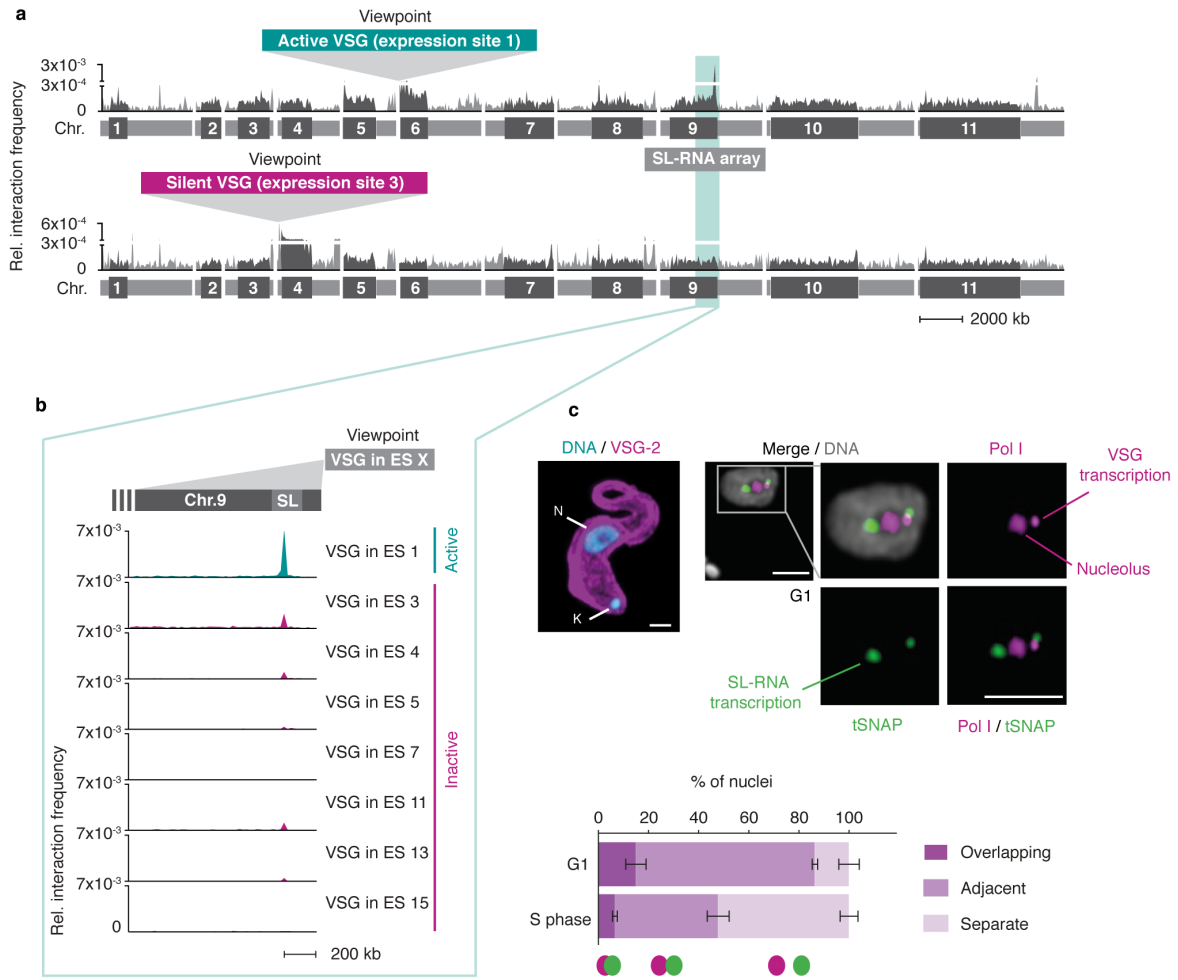


Fig. 1: The active VSG expression site (ES) stably interacts with the spliced leader RNA (SL) array. a, Hi-C (virtual 4C) analysis, viewpoints: active VSG gene in ES1 (*VSG-2*, top panel) and silent VSG gene in ES 3 (*VSG-6*, bottom panel). Relative interaction frequencies between the viewpoint and the 11 megabase chromosomes are shown. Chromosome cores, dark grey; sub-telomeric regions, light grey. The hemizygous sub-telomeric regions are displayed in the following order: 5´(haplotype A)–5´(haplotype B)–diploid chromosome core–3´(haplotype A)–3´(haplotype B). Bin size 50 kb. **b,** Virtual 4C analysis, viewpoints: active VSG gene in ES 1 and inactive VSG genes in ES 3, 4, 5, 7, 11, 13 and 15. Relative interaction frequencies between the viewpoint and the SL-RNA locus on the right arm of chr. 9 is plotted. Bin size 20 kb. The analyses in **a**-**b** are based on Hi-C experiments with *VSG-2* expressing cells (n=2, average interaction frequencies are shown). **c,** Immunofluorescence-based colocalisation studies of tSNAP^{myc} (SL-RNA locus marker – SL-RNA transcription compartment) and a nucleolar and active-*VSG* transcription compartment marker (Pol I, largest subunit) using super resolution microscopy. The stacked bar graph depicts proportions of G1 or S phase nuclei with overlapping, adjacent or separate signals for the SL-RNA and VSG transcription compartments. Values are averages of three independent experiments and representative of two independent biological replicates (≥100 G1 or S phase nuclei); error bars, SD. Detailed *n* and *p* values are provided in Data S1 sheet 3. DNA was counter-stained with DAPI; the images correspond to maximal 3D projections of stacks of 0.1 µm slices; scale bars 2 µm. N, nucleus; K, kinetoplast (mitochondrial genome).

To determine whether the interaction with the SL-RNA transcription compartment is specific for the active VSG gene, and therefore changes following a VSG switching event, we performed Hi-C experiments using an isogenic *T. brucei* cell line expressing a different VSG isoform, *VSG-13* (**Fig. 2a**) [19]. *VSG-13* resides within expression site 17, which is located on

one of five intermediate-sized chromosomes. The presence of co-transcribed resistance markers upstream of *VSG-2* in expression site 1 and *VSG-13* in expression site 17 allowed us to specifically select for parasites expressing *VSG-13* through drug selection (**Fig. 2a**). The exclusive activity of expression site 1 or 17 was verified by RNA-seq (**Fig. 2a**).

Hi-C analysis revealed that *VSG-2* - SL-RNA interactions dropped 20-fold to average inter-chromosomal interaction levels in parasites expressing *VSG-13*, while interactions between the newly activated *VSG-13* and the SL-RNA locus increased 36-fold (**Fig. 2b, left and middle panel**). We found that upon activation of each expression site, the bin harboring the respective active VSG gene displayed the strongest interaction with the SL-RNA locus, suggesting that the VSG gene itself, not its promoter, interacts with the splicing locus (**Fig. 2c**). In addition, we detected decreased interaction of the inactivated *VSG-2* gene with the transcribed chromosome cores, while the activated *VSG-13* gene displayed increased interaction with chromosome cores (**Extended Data Fig. 2a**). This observation indicates that activation of a VSG gene is intimately linked to a transition from a silent to an actively transcribed compartment within the nucleus. Conversely, VSG gene inactivation results in a transition from an active to an inactive nuclear compartment.

To further explore the relationship between SL-RNA interaction frequency and gene expression, we performed Hi-C analyses using insect stage parasites that do not express any VSGs, but instead express a different group of surface antigens called procyclin genes. Confirming the importance of the SL-RNA interaction, the GPEET and EP1 procyclin genes displayed increased interaction frequency with the SL-RNA locus upon activation in insect stage cells (**Fig. 2b, right panel, and Extended Data Fig. 2b**). Like VSG genes, procyclin genes are transcribed by RNA Pol I at high levels and require efficient *trans*-splicing for mRNA maturation. Thus, Hi-C analyses of *T. brucei* cell lines expressing different antigens indicate that interactions with the SL-RNA locus are dynamic and specific for the actively transcribed antigen coding genes.

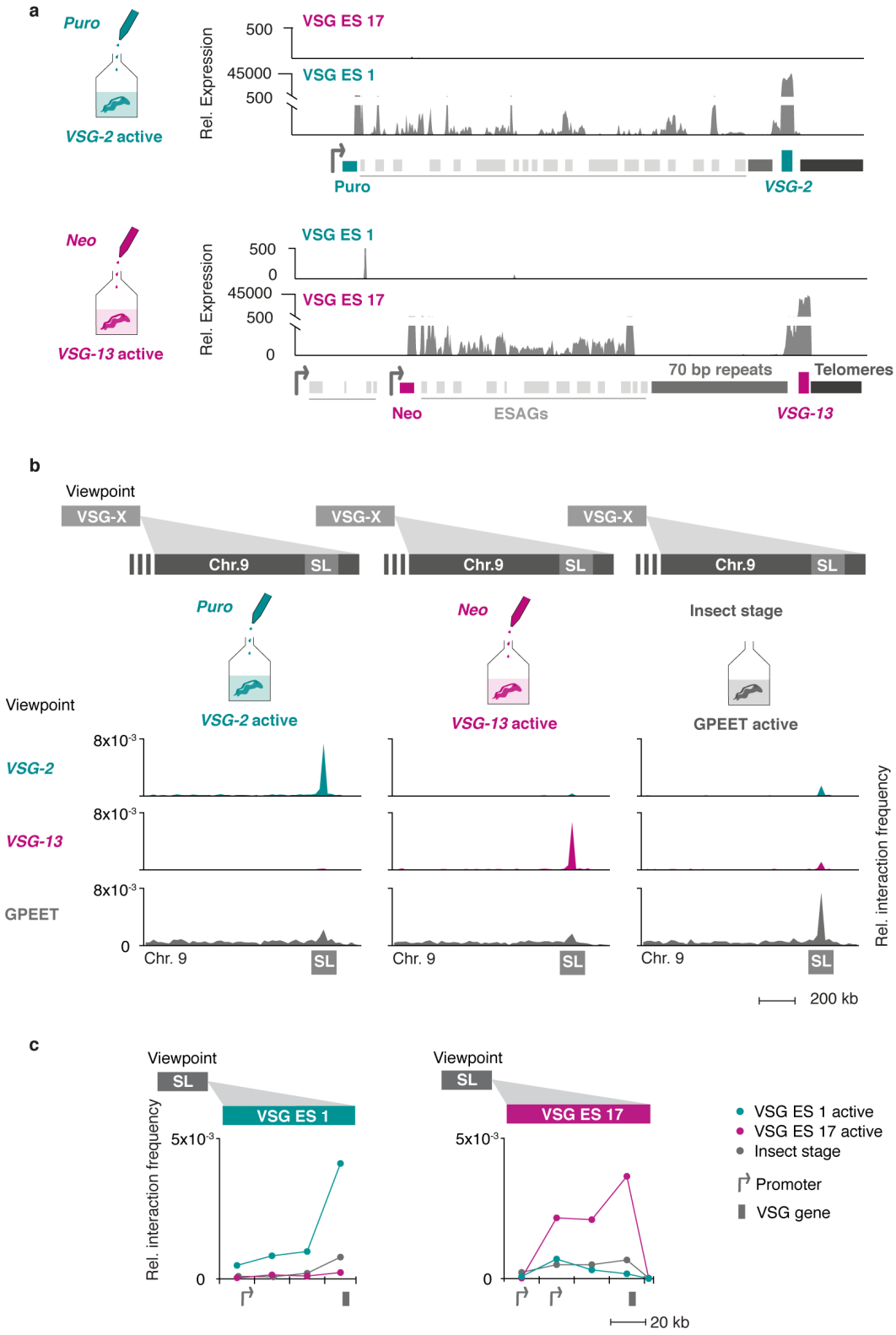


Fig. 2: The interaction between the active VSG gene and the SL-RNA locus is dynamic and changes during a switch in VSG expression. a, Transcriptome analyses of isogenic cell lines after selection with puromycin (active: *VSG-2*), top panel, or neomycin (active: *VSG-13*), bottom panel. For each condition, the average of three biological replicates is plotted. **b,** Hi-C (virtual 4C) analysis, viewpoints: *VSG-2*, *VSG-13*, GPEET gene array (chr. 6). Relative interaction frequencies between the viewpoint and the SL-RNA locus on chr. 9 are plotted. Bin size 20 kb. **c,** Hi-C (virtual 4C) analysis, viewpoint: SL-RNA locus (chr. 9). Relative interaction frequencies between the viewpoint and VSG ES 1 or ES 17 are shown. Bin size 20 kb. The analyses in **b** and **c** are based on Hi-C experiments with cells expressing *VSG-2*, *VSG-13* or insect stage cells expressing procyclin genes (n=2, the average is shown).

Previously, we had shown that the bipartite VEX-complex is associated with the actively transcribed VSG gene and maintains monogenic VSG expression but that VEX1 and VEX2 only partially overlap each other [16]. Given a similar juxtaposition of the VSG transcription and the SL-RNA transcription compartments, we sought to investigate the relationship between the VEX complex and these transcription compartments in more detail. Using optimized immunofluorescence staining protocols and super-resolution microscopy, we were able to detect two VEX1 foci in the majority of G1 cells (55 +/-4 %). These VEX1 signals specifically co-localized with the SL-RNA transcription compartments (**Fig. 3a**). In contrast, the majority of G1 cells (97 +/-1 %) only had one VEX2 focus, which specifically co-localized with the VSG transcription compartment (**Fig. 3b**). As expected, one VEX1 focus was adjacent to the VSG transcription compartment (**Extended Data Fig. 3a**) while the VEX2 focus was adjacent to one of the two SL-RNA transcription compartments (**Extended Data Fig. 3b**). Thus, our IFAs revealed association of VEX1 with the SL-RNA transcription compartments and VEX2 with the VSG transcription compartment. To verify a specific interaction between VEX1 and the SL-RNA locus, we reanalyzed published VEX1-ChIP-seq data, previously only mapped to VSG expression sites [16]. The ChIP-seq data revealed a striking enrichment of VEX1 at the SL-RNA locus that was greater than at any other gene, including the active VSG gene (**Fig. 3c, Extended Data Fig. 3c; Data S1, sheet 1**). These data suggest that the VEX-complex connects the VSG and SL-RNA transcription compartments.

Although VSG and SL-RNA transcription compartments separate during S phase (**Fig. 1c**), VEX1 does not separate from the SL-RNA transcription compartment and VEX2 does not separate from the VSG transcription compartment (**Fig. 3a and b**). Also, consistent with the loss of VSG expression in insect-stage cells, SL-RNA transcription compartments can still be identified through tSNAP localisation, while the VEX and VSG transcription compartments are reorganized and lost, respectively (**Extended Data Fig. 3d-e**). These results indicate that VEX2 marks the VSG transcription compartment in a developmental stage specific manner,

which in bloodstream stage cells may facilitate the re-establishment of compartment connectivity to propagate expression of a specific antigen. Consistent with this idea, VEX complex reassembly after DNA replication is dependent upon CAF-1 histone chaperone function [16].

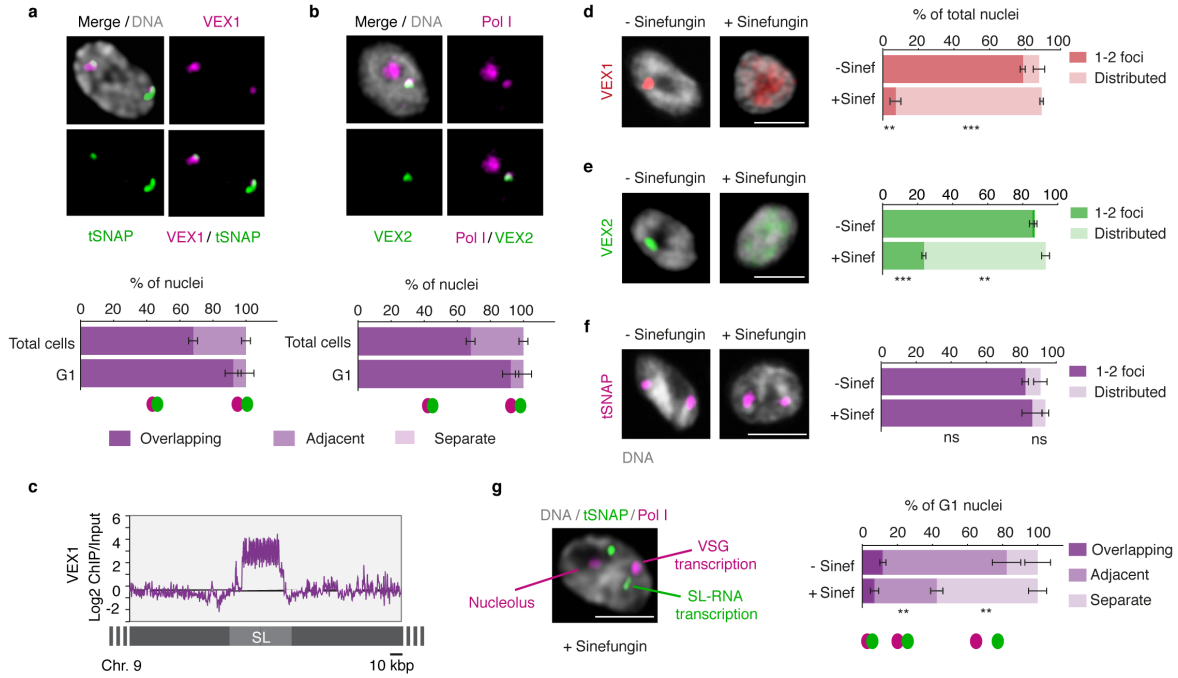


Fig. 3: The VEX complex associates with both the active-*VSG* and the Spliced Leader (SL)-locus in a splicing-dependent manner. a-b, Immunofluorescence-based colocalisation studies of VEX1^{myc} / tSNAP^{GFP} and ^{GFP}VEX2 / Pol I. tSNAP and Pol I are used as markers for the SL-RNA and VSG transcription compartments, respectively. The stacked bar graphs depict proportions of nuclei with overlapping, adjacent or separate signals: values are averages of four (**a**) or two (**b**) independent experiments (≥200 nuclei for total cell counts; ≥100 nuclei for G1 phase). **c,** VEX1^{myc} chromatin immunoprecipitation followed by next generation sequencing (ChIP-Seq) analysis. The graph depicts log2 fold change of ChIP signal versus input sample across the SL-RNA locus. Bin size 300 bp. **d-f,** Immunofluorescence analysis of VEX1^{myc} (**d**), ^{myc}VEX2 (**e**) and tSNAP^{myc} (**f**) before and after sinefungin treatment (5 µg ml^{-1} for 30 min at 37°C). Cells displaying no detectable signal (<10%) were excluded. Values are averages of two independent experiments (≥200 nuclei each). **g,** Immunofluorescence-based colocalisation studies of the SL-RNA transcription (tSNAP^{myc}) and the VSG transcription (Pol I, large subunit) compartments following treatment with sinefungin. The stacked bar graph depicts proportions of G1 nuclei with overlapping, adjacent or separate signals and values are averages of two independent experiments and two biological replicates (≥100 G1 nuclei). The studies in **a-b** / **d-g** were undertaken using super resolution microscopy and the images correspond to maximal 3D projections of stacks of 0.1 µm slices; DNA was counter-stained with DAPI; scale bars 2 µm. In **d-g,** a two-tailed paired Student's *t*-test was used to compare non-treated versus treated nuclei for each category: ns, non-significant; **, $p < 0.01$; ***, $p < 0.001$. Experiments in **a-b** / **d-g** are representative of at least two independent biological replicates and detailed *n* and *p* values are provided in Data S1 sheet 3.

Given the close spatial proximity between the site of VSG transcription and the site of SL-RNA transcription, we next questioned whether the splicing process itself impacts the connection between these compartments. We found that inhibition of *trans*-splicing with sinefungin [20] disrupted both VEX1 (**Fig. 3d**) and VEX2 (**Fig. 3e**) localization within 30 min,

while the tSNAP transcription factor was not affected under the same conditions (**Fig. 3f**). Notably, inhibition of splicing by sinefungin also disrupted the connection between the VSG and SL-RNA transcription compartments, revealed by separation of the Pol I and tSNAP signals (**Fig. 3g**); neither the VEX, nor the tSNAP protein levels were affected by sinefungin treatment (**Extended Data Fig.3f**). Thus, VEX protein localization and the juxtaposition of the VSG and the SL-RNA transcription compartments are dependent on mRNA splicing activity.

Next, we aimed to investigate the mechanism by which the VEX complex ensures monogenic VSG expression. Previously, we found that VEX2 depletion leads to a strong activation of VSG genes located in previously silent expression sites [16]. Thus, following our observation that the VEX-complex spans the VSG transcription and the SL-RNA transcription compartments, we sought to determine whether VEX2 functions as a connector or as an exclusion factor. That is, whether following VEX2 depletion the active VSG gene loses connectivity to the SL-RNA transcription compartment or whether previously inactive expression sites start to interact with the SL-RNA transcription compartment. To test these models, we analyzed the distance between the VSG transcription compartment and the SL-RNA transcription compartment following depletion of VEX-complex components (**Extended Data Fig. 4a-b**). IFA data revealed a disruption of compartment connectivity in 45% of G1 nuclei following 12 hours of VEX2 knockdown ($P < 0.001$) and 60% following VEX1 - VEX2 double-knockdown ($P < 0.01$) (**Fig. 4a, Extended Data Fig. 4c**); the mean sub-compartment distances were 60 nm in control cells and 167 nm following VEX-complex knockdown (**Fig. 4a-b, Extended Data Fig. 4d**). At later time points after the induced VEX2 knock down, the RNA Pol I signal disperses [18], indicating that VEX2 sustains a local reservoir of Pol I at the active VSG gene. Compartment connectivity was not significantly disrupted following VEX1-knockdown.

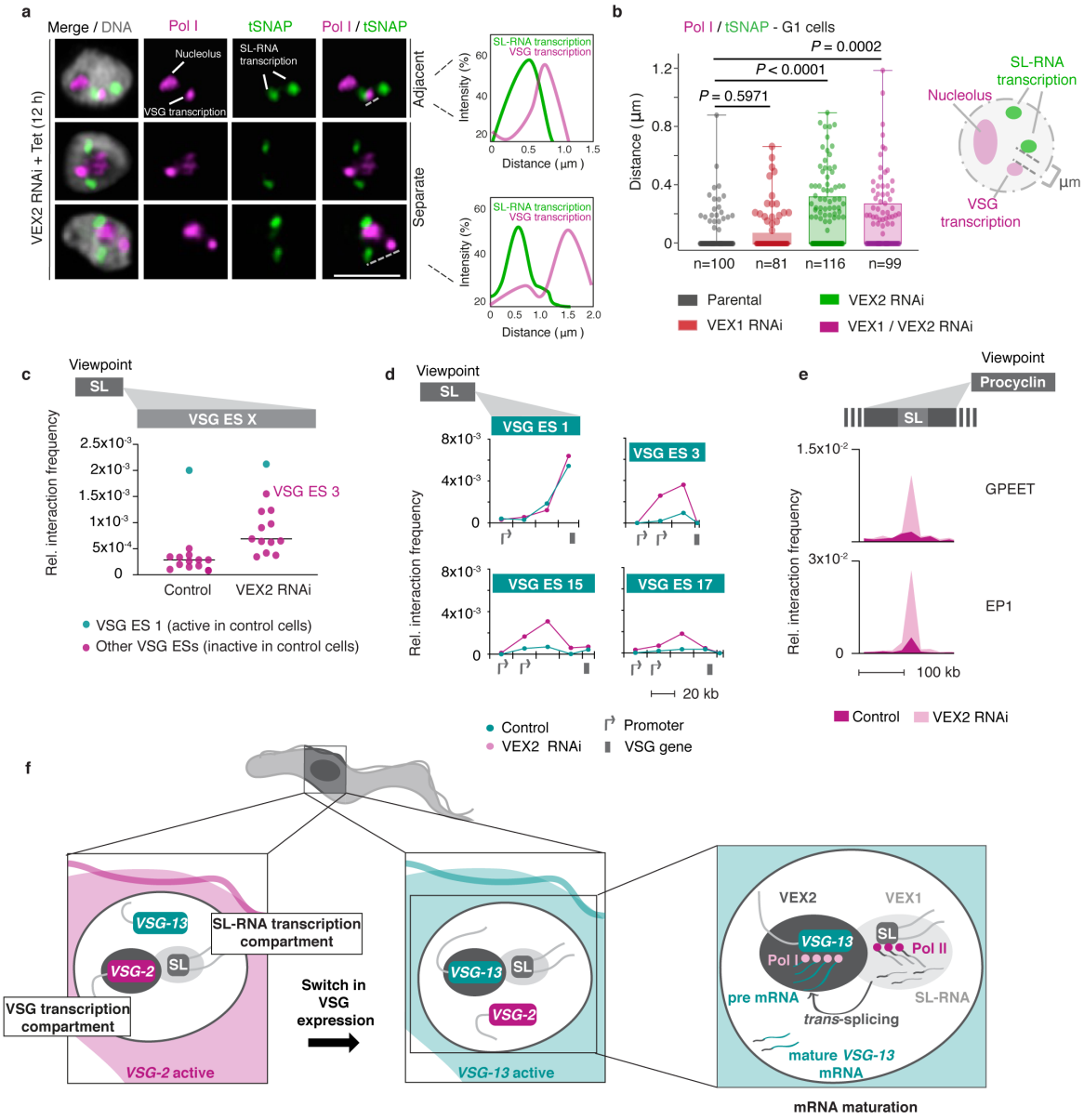


Fig. 4: The exclusive association between the active VSG and the *SL*-locus is VEX2-dependent. a-b, Immunofluorescence and super resolution microscopy based colocalisation studies of tSNAP^{myc} (SL-RNA transcription compartment) and Pol I (nucleolus and active-VSG transcription compartment) following VEX1, VEX2 and VEX1/VEX2 knockdown. **a,** Representative images, all nuclei are G1. On the right-hand side, two representative histograms depict the distribution of signal intensity across the distance indicated by the dashed lines. Images correspond to maximal 3D projections of stacks of 0.1 µm slices; DNA was counter-stained with DAPI; scale bars 2 µm. **b,** The box plot depicts the distance between the active-VSG and the nearest SL-RNA transcription compartment (≥ 81 G1 nuclei) following VEX1, VEX2 and VEX1/VEX2 knockdown; the centre lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend from maximal to the minimal values; all data points are shown. The distances for each knockdown condition were compared to the parental cell line using a two-tailed unpaired Student's *t*-test. **c,** Hi-C (virtual 4C) analysis, viewpoint: SL-RNA locus (chr. 9). Relative interaction frequencies between the viewpoint and the VSG expression sites are shown before and after VEX2 knockdown. Each dot represents the average value for one expression site. Bin size 20 kb. **d,** Virtual 4C analysis, viewpoint: SL-RNA locus (chr. 9). Relative interaction frequencies between the viewpoint and VSG expression sites is shown. Bin size 20 kb. **e,** Virtual 4C analysis, viewpoint: EP1 (chr. 10) or GPEET gene array (chr. 6). Relative interaction frequencies between the viewpoint and the SL-RNA locus are plotted. Bin size 20 kb. The analyses in **c** to **e** are based on Hi-C experiments with cells before and 24 h after VEX2 knockdown (n=3, the average is shown). **f,** Schematic model for monogenic VSG expression. A strong inter-chromosomal interaction between the *SL*-array and the active VSG gene facilitates spatial integration of transcription and mRNA maturation. VEX1 and VEX2 are primarily *SL*- and active-VSG associated, respectively, and sustain monogenic VSG expression by excluding other VSGs. The VSG-SL organelle is reconfigured upon activation of a different VSG.

To explore the role of VEX2 in controlling interactions between antigencoding genes and SL-RNA loci, we performed Hi-C analyses in VEX2-depleted cells. After 24 hours of VEX2 depletion, all previously silent expression sites displayed increased interaction frequencies with the SL-RNA locus. (**Fig. 4c**). The interaction between VSG expression site 3 and the SL-RNA locus showed the strongest increase. Notably, this is the expression site containing the most de-repressed VSG (*VSG-6*) following VEX2 depletion [16]. Interaction frequencies of the active VSG expression site 1 with the SL-RNA locus remained unchanged, correlating with sustained and dominant *VSG-2* expression. Overall, SL-RNA interactions correlate with VSG transcript levels before and after VEX2 knockdown (**Extended Data Fig. 5a**).

Besides the VSG genes located in previously silent expression sites, expression site associated genes were also strongly upregulated following VEX2 knockdown [16]. In line with this finding, we observed the largest increase in SL-RNA interactions for the regions upstream of the VSG gene in each de-repressed expression site, where expression site associated genes are located (**Fig. 4d, Extended Data Fig. 5b**). Thus, VEX2 restricts interactions between silent VSG expression sites, the expression site associated genes in particular, and the SL-RNA locus.

As a third group of RNA Pol I transcribed genes, insect stage specific procyclin genes are upregulated upon VEX2 depletion [16]. Correlating with these data, following VEX2-knockdown, we found GPEET and EP1 procyclin genes to exhibit strongly increased interaction frequencies with the SL-RNA array and also with VSG expression sites (**Fig. 4e, Extended Data Fig. 5c**). Thus, our data suggest that VEX2 may have a dual function: specifically enhancing mRNA splicing of the VSG gene that is connected to the SL-RNA transcription compartment and, at the same time, excluding all other VSG expression sites and procyclin genes from the SL-RNA compartment to ensure monogenic VSG expression.

By combining proximity ligation and super-resolution microscopy, we were able to demonstrate spatial integration of the active VSG expression site and a genomic locus

important for RNA maturation. Our data show that this supramolecular assembly is composed of a VSG transcription compartment with the active VSG gene, RNA Pol I and VEX2 and an SL-RNA transcription compartment with the SL-RNA array, RNA Pol II, VEX1 and the tSNAP complex, presumably together with other factors important for mRNA *trans*-splicing [18]. Based on these findings we propose a model in which VSG choice is intimately associated with an inter-chromosomal interaction, bringing together two nuclear compartments to ensure efficient VSG mRNA processing at only one expression site (**Fig. 4f**). In the VSG transcription compartment, the VSG gene is transcribed by highly processive RNA Pol I, generating large amounts of VSG pre-mRNA that requires efficient processing to prevent premature degradation. In the SL-RNA transcription compartment, SL-RNA, an essential substrate for maturation of every mRNA, is produced. The close spatial proximity of the two compartments in a single locus provides a sufficiently high concentration of *trans*-splicing substrate to ensure the efficient maturation of highly abundant VSG transcripts. According to this model, VEX2 is the key molecule bridging the two compartments and excluding all but one VSG expression site from the SL-RNA transcription compartment, thereby ensuring expression of a single VSG gene per cell.

RNA processing as the limiting factor in monogenic VSG expression has been proposed previously, based on the observations that all VSG expression site promoters are active, yet processive polycistronic transcription and mature transcripts are specific to the single active expression site [6,7]. Given the observation that co-transcriptional RNA processing can affect elongation rates in other organism [21], it will be interesting to determine whether efficient VSG mRNA processing will exert a positive feedback on transcription elongation along expression sites. Also, it was shown recently that splicing can activate transcription [22]. It remains to be shown if factors located in the SL-RNA transcription compartment recruit the transcription machinery to the interacting VSG expression site and thereby enhance transcription. Notably, we also find other highly expressed housekeeping genes, such as core

histones or tubulin, to associate with the SL-RNA locus (**Extended Data Fig. 6a**) and with VEX1 (**Extended Data Fig. 6b**). Thus, interactions with the SL-RNA locus may play a broader role for the regulation of gene expression in *T. brucei*.

While the importance of intra-chromosomal interactions in regulating gene expression has been shown in many complex eukaryotes, the significance of inter-chromosomal interactions has been questioned. Interactions between different chromosomes were thought to require complicated, possibly error-prone mechanisms to be re-established following mitosis. Yet, our data demonstrate a stable interaction between the active VSG gene, located either on chr. 6 or on an intermediate-sized chromosome, and the SL-RNA array locus on chr. 9. The interaction is also stably propagated during cell-division; despite being resolved during S phase, the interaction is re-established after replication. To our knowledge, these findings represent the first report of a selective inter-chromosomal interaction that connects transcription and mRNA splicing.

Protein condensates have recently emerged as important features that compartmentalize nuclear functions; transcription control by RNA polymerases, for example [23]. Regulated switching between adjacent transcriptional and splicing condensates has been described in mammalian cells [24] and RNAs are major actors in facilitating genomic interactions and phase transitions [25]. Given the recent finding that members of a family of helicases are global regulators of RNA-containing, phase-separated organelles [26], it is tempting to speculate that maintenance of VSG transcription-maturation compartment connectivity is similarly regulated by the putative RNA helicase VEX2 [16]. We show here that in *T. brucei*, the assembly of two membrane-less nuclear condensates, each with a specific function in VSG gene transcription and RNA maturation, and both containing protein and RNA molecules [12,18], regulates monogenic VSG expression. By shaping a highly selective and specific genome architecture, VEX2 allows only one VSG gene to productively interact with the mRNA splicing compartment.

Acknowledgements: We thank the Dundee Imaging Facility and J. Rouse for access to the Zeiss 880 Airyscan and Leica Confocal SP8 Hyvolution microscope, respectively, and S. Alsford (London School of Hygiene & Tropical Medicine) for the SNAP42 tagging construct. We further thank R. Cosentino and all members of the Siegel, Ladurner, Meissner and Boshart labs for valuable discussion, T. Straub (Core facility Bioinformatics, BMC) for providing server space and help with the data analysis, the Core Unit Systems Medicine, University of Würzburg for NGS.

Funding: The work was funded by a Wellcome Trust Investigator Award to D.H. [100320/Z/12/Z], by the German Research Foundation [SI 1610/3-1], the Center for Integrative Protein Science (CIPSM) and by an ERC Starting Grant [3D_Tryps 715466]. The University of Dundee Imaging Facility is supported by the MRC Next Generation Optical Microscopy award [MR/K015869/1]. L.S.M.M. was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Science, University of Würzburg.

Author contributions: Experiments were designed by J.F., V.L., L.S.M.M., D.H. and T.N.S. and carried out by J.F., V.L. and L.S.M.M., unless indicated otherwise. Initial IFA experiments were performed by L.G.. RNAi, chemical inhibition and super resolution and other microscopy experiments were performed and analysed by J.F.. Hi-C experiments were performed by V.L. and L.S.M.M. and computational analysis was carried out by B.G.B., V.L. and L.S.M.M.. RNA-seq experiments were performed by V.L.; data analysis was carried out by V.L. and B.G.B.. ChIP-seq data analysis was carried out by S.H.. Funding was acquired by D.H. and T.N.S.. The work was supervised by D.H. and T.N.S.. The manuscript was written by J.F., V.L., D.H. and T.N.S. and edited by all other co-authors.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: Code and high-throughput sequencing data generated for this study have been deposited at GitHub (https://github.com/bgbrink/PRJEB35632) and in the European Nucleotide Archive (ENA) under primary accession number PRJEB35632, respectively. ChIP-seq data have also been deposited in the ENA (accession no. PRJEB25352). Processed data and results are available under https://doi.org/10.5281/zenodo.3628213.

References:

1 Bashkirova, E. & Lomvardas, S. Olfactory receptor genes make the case for inter-chromosomal interactions. *Curr Opin Genet Dev* **55**, 106-113 (2019).

2 Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* **7**, 493-503 (2009).

3 Aresta-Branco, F., Erben, E., Papavasiliou, F. N. & Stebbins, C. E. Mechanistic similarities between antigenic variation and antibody diversification during *Trypanosoma brucei* infection. *Trends Parasitol* **35**, 302-315 (2019).

4 Horn, D. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol* **195**, 123-129 (2014).

5 Muller, L. S. M. *et al.* Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**, 121-125 (2018).

6 Kassem, A., Pays, E. & Vanhamme, L. Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* **111**, 8943-8948 (2014).

7 Vanhamme, L. *et al.* Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. *Mol Microbiol* **36**, 328-340 (2000).

8 Michaeli, S. *Trans*-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol* **6**, 459-474 (2011).

9 Bertero, A. *et al.* Dynamics of genome reorganization during human cardiogenesis reveal an RBM20-dependent splicing factory. *Nat Commun* **10**, 1538 (2019).

10 Khanna, N., Hu, Y. & Belmont, A. S. HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr Biol* **24**, 1138-1144 (2014).

11 Navarro, M. & Gull, K. A pol I transcriptional body associated with *VSG* mono-allelic expression in *Trypanosoma brucei*. *Nature* **414**, 759-763 (2001).

12 Budzak, J. *et al.* Dynamic colocalization of 2 simultaneously active *VSG* expression sites within a single expression-site body in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* **116**, 16561-16570 (2019).

13 Chaves, I. *et al.* Subnuclear localization of the active variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* **95**, 12328-123338 (1998).

14 Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat Rev Genet* **20**, 437-455 (2019).

15 Glover, L., Hutchinson, S., Alsford, S. & Horn, D. VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. *Proc Natl Acad Sci U S A* **113**, 7225-7230 (2016).

16 Faria, J. *et al.* Monoallelic expression and epigenetic inheritance sustained by a *Trypanosoma brucei* variant surface glycoprotein exclusion complex. *Nat Commun* **10**, 3023 (2019).

17 Zheng, Y., Ay, F. & Keles, S. Generative modeling of multi-mapping reads with mHi-C advances analysis of Hi-C studies. *Elife* **8** (2019).

18 Hury, A., Goldshmidt, H., Tkacz, I. D. & Michaeli, S. Trypanosome spliced-leader-associated RNA (SLA1) localization and implications for spliced-leader RNA biogenesis. *Eukaryot Cell* **8**, 56-68 (2009).

19 Figueiredo, L. M., Janzen, C. J. & Cross, G. A. A histone methyltransferase modulates antigenic variation in African trypanosomes. *PLoS Biol* **6**, e161 (2008).

20 McNally, K. P. & Agabian, N. *Trypanosoma brucei* spliced-leader RNA methylations are required for *trans* splicing *in vivo*. *Mol Cell Biol* **12**, 4844-4851 (1992).

21 Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J. & Nogues, G. Multiple links between transcription and splicing. *Rna* **10**, 1489-1498 (2004).

22 Fiszbein, A., Krick, K. S., Begg, B. E. & Burge, C. B. Exon-mediated activation of transcription starts. *Cell* **179**, 1551-1565 (2019).

23 Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13-23 (2017).

24 Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543-548 (2019).

25 Li, X. & Fu, X. D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet* **20**, 503-519 (2019).

26 Hondele, M. *et al.* DEAD-box ATPases are global regulators of phase-separated organelles. *Nature* **573**, 144-148 (2019).

491 **Methods:**

492 No statistical methods were used to predetermine sample size. The experiments were not
493 randomized and investigators were not blinded to allocation during experiments and outcome
494 assessment.

495

496 ***T. brucei* growth and manipulation.** Bloodstream-form *T. brucei*, Lister 427 and 2T1 cells [27],
497 both wild-type with respect to VEX1, VEX2 and SNAP42 subunits, were grown in HMI-11
498 medium and genetically manipulated using electroporation [28]; cytomix was used for all
499 transfections. Puromycin, phleomycin, hygromycin and blasticidin were used at 2, 2, 2.5 and
500 10 μg ml^{-1} for selection of recombinant clones; and at 1, 1, 1 and 2 μg ml^{-1} for maintaining
501 those clones, respectively. RNAi experiments were undertaken through tetracycline induction
502 at 1 μg ml^{-1}. A double selection *T. brucei* cell line was used that derived from the Lister 427
503 bloodstream-form MITat 1.2 isolate [19]. A neomycin resistance gene in VSG expression site 17
504 and a puromycin resistance gene in VSG expression site 1 allowed the selection for a
505 homogenous cell population that either expressed *VSG-2* from expression site 1 or *VSG-13*
506 from expression site 17. Cells were cultivated with either 10 μg ml^{-1} of neomycin (also referred
507 to as N50 cells) or 0.1 μg ml^{-1} of puromycin (referred to as P10 cells). Established procyclic-
508 form *T. brucei*, Lister 427 cells were grown in SDM-79 at 27 °C and genetically manipulated
509 using electroporation as above. Blasticidin or hygromycin were used at 10 or 50 and 2 or 1 µg
510 ml^{-1} for selection and maintenance, respectively.

511

512 **Plasmids.** The VEX1 (Tb927.11.16920, 574 bp) [15], VEX2 (Tb927.11.13380, 471 bp) [16] and
513 VEX1/VEX2 [16] RNAi cassettes were excised prior to electroporation by digesting with *Asc*I.
514 The VEX1^{12myc} [15] and SNAP42^{12myc} [29] *C*-terminal tagging vectors were linearised with *Sph*I.
515 The ^{6myc}VEX2 and ^{GFP}VEX2 *N*-terminal tagging vectors were linearised with *Xho*I. The
516 SNAP42 GFP *C*-terminal tagging vector was made by replacing the 12 x c-myc tag and as

517 also linearised with *Sph*I, respectively. Linearised RNAi constructs, under the control of
518 tetracycline-inducible promoters, were transfected into 2T1 cells, which allow for targeting to
519 a single genomic locus validated for robust inducible expression [27].
520
521 **ChIP-seq.** ChIP-Seq was carried out as described in [16]. Reads were aligned to the 11 curated
522 megabase chromosomes from the TREU927 strain genome sequence [30], and a non-
523 redundant set of BES and mVSG contigs from the Lister 427 strain [31-33] using bowtie2 [34] in
524 very-sensitive alignment mode, and alignments were compressed and sorted using
525 samtools [35]. Bowtie2 attempted to align 54.0 and 49.9 million read pairs with 70.84 and 82.43
526 % success rates, respectively, resulting in 38.3 and 41.1 million aligned read pairs. PCR
527 duplicate reads were removed using Picard MarkDuplicates
528 (http://broadinstitute.github.io/picard/) resulting in 26.8 and 41.3 million aligned read pairs for
529 analysis. Alignments were visually inspected with the Artemis genome browser [36]. Circular
530 plots (**Extended Data Fig. 3c**) were generated using the R library circlize [37] and bedgraph
531 files for log2 fold change (**Fig. 3c, Extended Data Fig. 3c and Extended Data Fig. 6b**) were
532 generated using deeptools2 [38]. Bedgraphs were generated with 1kb bins and the option
533 smoothLength 5000. Spliced leader RNA sequences were annotated using the sequences:
534 Promoter:CGTTTCTGGCACGACAGTAAAATATGGCAAGTGTCTCAAAACTGCCTGTACA
535 GCTTATTTTTGGGACACACCCATGCTTTC…Transcript…AACTAACGCTATTATTAGAACA
536 GTTTCTGTACTATATTGGTATGAGAAGCTCCCAGTAGCAGCTGGGCCAACACACGCATT
537 TGTGCTGTTGGTTCCCGCCGCATACTGCGGGAATCTGGAAGGTGGGGTCGGATGACCT
538 C and the 'transcript' features were plotted with TSS and TES denoting the 5' and 3'
539 extremities. Fold enrichment traces covering the spliced leader locus were calculated directly
540 using deeptools bamCompare. Heat maps (**Extended Data Fig. 3c and Extended Data Fig.**
541 **6b**) were generated using deeptools2 bamCompare, computeMatrix and plotHeatmap [38]
542 and resulting vector graphics files were then assembled into figures using Adobe

Illustrator. Genomic regions for tandem genes and arbitrarily selected genes genes with a paralog count of 0 were assembled in bed files, using annotated mRNA sequences from TriTrypDB v5.1 of the TREU927 genome sequence. All scripts necessary to reproduce the ChIP-Seq analyses have been deposited together with the results of those analyses under https://doi.org/10.5281/zenodo.3628213.

Protein blotting. Protein samples were run according to standard protein separation procedures, using SDS-PAGE. However, for VEX2 detection, the use of Bis-Tris gels with a neutral pH environment and a Bis-Tris/Bicine based transfer buffer (containing a reducing agent and 10% methanol) were critical for protein separation and transfer, respectively (NuPAGE, Invitrogen). Otherwise, western blotting was carried out according to standard protocols. The following primary antibodies were used: rabbit α-VEX2 (1:1,000), rabbit α-pol-I largest subunit [15] (1:500), rabbit α-VSG-2 (1:20,000), rabbit α-VSG-6 (1:20,000), mouse α-c-myc (Millipore, clone 4A6, 1:7,000), rabbit α-GFP (Abcam Ab290, 1:1,000) and mouse α-EF1α (Millipore, clone CBP-KK1, 1:20,000). We used horseradish peroxidase coupled secondary antibodies (α-mouse and α-rabbit, Biorad, 1:2,000). Blots were developed using an enhanced chemiluminescence kit (Amersham) according to the manufacturer's instructions. Densitometry was performed using Fiji v. 2.0.0.

Microscopy. Immunofluorescence microscopy was carried out according to standard protocols. For wide field microscopy (**Extended Data Fig. 4a**), the cells were attached to 12-well 5 mm slides (Thermo Scientific). For super resolution microscopy, the cells were attached to poly-L-lysine treated coverslips (thickness $1^{1/2}$ mm), stained and only then mounted onto glass slides. For colocalisation studies with Pol I we used antigen-retrieval. Prior to permeabilization, fixed cells were rehydrated in PBS for 5 min at RT, held at 95 °C for 60 s in freshly prepared antigen retrieval buffer (100 mM Tris, 5% urea, pH 9.5) and then washed 3 x

5 min in PBS at RT. Cells were mounted in Vectashield with DAPI (wide field) or stained with 1 μg ml^{-1} DAPI for 10 min and then mounted in Vectashield without DAPI (super resolution). In *T. brucei*, DAPI-stained nuclear and mitochondrial DNA were used as cytological markers for cell-cycle stage; one nucleus and one kinetoplast (1N:1K) indicates G1, one nucleus and an elongated kinetoplast (1N:eK) indicates S phase, one nucleus and two kinetoplasts (1N:2K) indicates G2/M and two nuclei and two kinetoplasts (2N:2K) indicates post-mitosis [39,40]. Primary antisera were rat α-VSG-2 (1:10,000), rabbit α-VSG-6 (1:10,000), rabbit α-GFP (Invitrogen, 1:250; Abcam, 1:500), mouse α-myc (New England Biolabs, clone 9B11, 1:2,000), rabbit α-pol-I largest subunit [15] (1:100) or rabbit α-NOG1 [41] (1:500). The secondary antibodies were Alexa Fluor conjugated goat antibodies (Thermo Scientific): α-mouse, α-rat and α-rabbit, AlexaFluor 488 or Alexa Fluor 568 (1:1,000 or 1:2,000, for super resolution or wide field field microscopy, respectively). For wide field microscopy, cells were analysed using a Zeiss Axiovert 200M microscope with an AxioCam MRm camera and the ZEN Pro software (Carl Zeiss, Germany). The images were acquired as z-stacks (0.1-0.2 μm) and further deconvolved using the fast iterative algorithm in Zen Pro. For super resolution microscopy, cells were analysed using a Leica TCS SP8 confocal laser scanning microscope in Hyvolution Mode and the Leica Application Suite X (LASX) software (Leica, Germany) or a Zeiss 880 Airyscan and the Zeiss ZEN software (Carl Zeiss, Germany). The Hyvolution mode allows super resolution level images, general used settings: highest resolution / lowest speed; pinhole 0.5. All the super resolution images correspond to maximal 3D projections by brightest intensity of stacks of approximately 30 slices of 0.1 μm - Images with DNA in grey. For all quantifications, images were acquired with the same settings and equally processed. All the images were processed and scored using Fiji v. 2.0.0. [42]. Sinefungin was applied at 2 μg ml^{-1} for 30 minutes at 37ºC. VEX1, VEX2 and tSNAP foci and Pol I nucleolar and ESB signals could be detected in over 85-90% of nuclei. For all experiments involving RNAi, the knockdown was always verified by Western-Blot and the VSG dereppression phenotype confirmed by IFA and/or FACS analysis.

595 Counts in total cells or specific cell cycle phases were performed in >200 or >100 nuclei,
596 respectively. All quantifications are averages or representative of at least two biological
597 replicates and independent experiments. For the ESB / tSNAP localisation following VEX2 or
598 VEX1 / VEX2 RNAi in **Fig. 4a-b and Extended Data Fig. 4c-d**, all the imaging and analysis
599 was performed at 12 h post induction, a timepoint where there was sufficient VEX2 knockdown
600 (**Extended Data Fig. 4b**) but both nucleolar Pol I and the ESB could be detected in > 85% of
601 cells; the ESB is not detectable at later time points [16]. Regarding the distance measurements
602 between the ESB and tSNAP compartment (**Fig. 4c**), a control measurement (**Extended Data**
603 **Fig. 4d**) was performed to make sure that the increase in the distance between the two protein
604 condensates following VEX2 or VEX1 / VEX2 RNAi was not a mere consequence of a
605 decrease in the ESB focus diameter. Moreover, the ESB / tSNAP localisation analyses
606 following VEX RNAi or sinefungin treatment were restricted to G1 cells to exclude any cell
607 cycle bias, as these protein condensates can separate during S phase (**Fig. 1c**).
608
609 **RNA-Seq.** RNA isolation: The RNA-seq experiment and data analysis was performed as
610 described previously [43], using three replicates each for *VSG-2* and *VSG-13* expressing cells.
611 45 million cells were harvested per replicate at 1,500 x g and 4 °C for 10 min. Cells were
612 washed with 1× TDB (5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20 mM Na_2HPO_4, 2 mM
613 NaH_2PO_4, 20 mM glucose pH 7.4). RNA isolation was performed using the NucleoSpin RNA
614 kit (Macherey-Nagel; cat. no. 740955.10) according to the manufacturer´s instructions with
615 minor changes. 3.8 µl of 1 M RNAse-free dithiothreitol (Sigma-Aldrich; cat. no. 10197777001)
616 and 1 µl of 1:10 Ambion ERCC RNA Spike-In Mix (ThermoFisherScientific; cat. no. 4456739)
617 was added to the cell lysis buffer prior to use. Removal of ribosomal RNA: rRNA was removed
618 by hybridization as described previously [43]. All solutions were kept free from nucleases. For
619 each hybridization reaction, 2 µg of total RNA was mixed with 10 µl of formamide
620 (SigmaAldrich; cat. no. F9037-100ML), 2.5 µl of 20× SSC (3 M NaCl, 0.3 M sodium citrate, the

pH was adjusted to 7.0 with HCl), 5 µl of 0.005 M EDTA pH 8 (stock solution 0.5 M; ThermoFisherScientific; cat. no. AM9260G), 2.48 µl of 100 µM rRNA depletion mix (total 4 µg of oligos) and RNAse-free water (ThermoFisherScientific; cat. no. AM9938) to a total volume of 50 µl. Hybridization was performed for 5 min at 80 °C, ramp down to 25 °C at intervals of 5 °C per minute. Subsequently, 2 µl of RNAse-OUT (ThermoFisherScientific; cat. no. 10777019) and 50 µl of 1x SCC containing 20% formamide were added. Dynabead MyOne Streptavidin C1 beads (ThermoFisherScientific; cat. no. 65001) were prepared as recommended by the manufacturer for RNA applications and immobilization of nucleic acids. Three rounds of oligo capture were performed, using 120 µl (1.2 mg) of magnetics beads per round. The resulting supernatant, containing rRNA-depleted RNA, was purified using RNeasy MinElute CleanUp Kit (QIAGEN; cat. no. 74204). Depletion of rRNAs was evaluated on a 1.2% TBE-agarose gel. cDNA synthesis, library preparation and sequencing: Synthesis of cDNA was performed using NEBNext Ultra Directional RNA Library Prep Kit from Illumina (New England Biolabs; cat. no. E7420) according to the manufacturer's instruction. The concentration of cDNA was measured using Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854) and a Qubit 2.0 Fluorometer (Invitrogen; cat. no. Q32866). To generate strand-specific RNA-seq libraries, uracil excision and removing of the second strand was performed prior to conversion of Y-shaped adapters. Therefore, 3 µl of USER enzyme (New England Biolabs; cat. no. M5505) were mixed with 16 µl of adapter-ligated DNA, 1 µl of TruSeq PCR primer cocktail (50 µM) and 20 µl of KAPA HiFi HotStart ReadyMix (KAPA Biosystems, cat. no. KK2601). USER digestion was performed at 37 °C for 15 min, followed by the published amplification protocol. Library concentrations were determined in duplicate using Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854) and a Qubit 2.0 Fluorometer (Invitrogen, cat. no. Q32866) and quantified using the KAPA Library Quantification Kit (KAPA Biosystems, cat. no. KK4824) according to the manufacturer's instruction. Strand-specific RNA-sequencing libraries were sequenced in paired-end mode on an Illumina NextSeq 500 sequencer (2×75 cycles). Processing of sequencing data: The

sequencing datasets were mapped to the TbruceiLister427_2018 genome assembly (release 43, downloaded from TriTrypDB [44]) using BWA-mem [45]. The alignments were converted from SAM to BAM format, sorted, indexed and filtered by alignment quality (q>0) using SAMtools version 1.9 [35]. To visualize read coverage, the number of reads was normalized per billion mapped reads and coverage files were generated in the wiggle format using COVERnant version 0.3.0 with the subcommand *ratio* [46].

***In situ* Hi-C.** In situ Hi-C was performed as previously described [5]. 2×10^8 cells were collected per replicate and resuspended in 40 ml of 1× trypanosome dilution buffer (1× TDB; 0.005 M KCl, 0.08 M NaCl, 0.001 M $MgSO_4 \times 7H_2O$, 0.02 M Na_2HPO_4, 0.002 M $NaH_2PO_4 \times 2H_2O$, 0.02 M glucose) or 1× PBS (insect stage cells). Cells were fixed in the presence of 1% formaldehyde for 20 min at room temperature by addition of 4 ml of 11% formaldehyde solution (50 mM Hepes-KOH pH 7.5, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 11% formaldehyde). The reaction was stopped by addition of 3 ml of 2 M glycine and incubation for 5 min at room temperature and 15 min on ice. Cells were washed twice in 1× TDB for bloodstream form cells or 1× PBS for insect stage cells, respectively, and the cell pellet was snap-frozen in liquid nitrogen. Cells were resuspended in 1 ml of permeabilization buffer (100 mM KCl, 10 mM Tris pH 8.0, 25 mM EDTA) supplemented with protease inhibitors (1.5 mM pepstatin A, 4.25 mM leupeptin, 1.06 mM PMSF, 1.06 mM TLCK) and digitonin (200 µM final concentration) and incubated for 5 min at room temperature. Cells were washed twice in 1× NEBuffer3.1 (NEB, B7003S) and resuspended in 342 µl of 1× NEBuffer3.1. After addition of 38 µl of 1% SDS, and an incubation at 65 °C for 10 min, SDS was quenched by addition of 43 µl of 10% Triton-X 100 (Sigma). Incubation was continued at room temperature for 15 min. Another 35 µl of water, 13 µl of 10× NEBuffer3.1 and 100 units of MboI (NEB, R0147M) were added and the chromatin was digested at 37 °C overnight while shaking. To inactivate MboI, the sample was incubated at 65 °C for 20 min. Restriction fragments were biotinylated by supplementing the reaction with 60 µl of fill-in mix (0.25 mM biotin-14-dATP (Life

673 Technologies, 19524016), 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP (Fermentas), 40
674 U of DNA polymerase I, large (Klenow) fragment (NEB, M0210)) and incubation at 23 °C for
675 4 h. The end-repaired chromatin was transferred to 665 µl of ligation mix (1.8% Triton-X 100,
676 0.18 mg BSA, 1.8× T4 DNA Ligase Buffer (Invitrogen, 46300018) and 5 µl of T4 DNA ligase
677 (invitrogen, 15224025) were added. The ligation was performed for 4 h at 16 °C with interval
678 shake. Crosslinks were reversed by adding 50 µl of 10 mg/ml proteinase K (65 °C for 4 h)
679 following addition of another 50 µl of 10mg/ml proteinase K, 80 µl of 5 M NaCl and 70 µl of
680 10% SDS (65 °C, overnight). DNA was precipitated with ethanol and resuspended in 257 µl
681 of TLE (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). SDS was added to a final concentration of
682 0.1% and the sample was split among two tubes for sonication (Covaris S220; microtubes,
683 175 W peak incident power, 10% duty factor, 200 cycles per burst, 240 s treatment). The
684 samples were recombined and the volume was adjusted to 300 µl with TLE. Fragments
685 between 100 and 400 bp in size were selected using Agencourt AMPure XP beads (Beckman
686 Coulter), according to the manufacturer's instructions. The DNA fragments were eluted off the
687 beads in 55 µl of TLE. For end-repair and biotin removal from un-ligated ends, 70 µl of end-
688 repair mix was added (1× Ligation buffer (NEB), 357 µM dNTPs, 25U T4 PNK (NEB, M0201),
689 7.5U T4 DNA polymerase I (NEB, M0203), 2.5U DNA polymerase I, large (Klenow) fragment
690 (NEB, M0210)) and incubated for 30 min at 20 °C and 20 min at 75 °C. To inactivate the
691 enzymes, EDTA was added to a final concentration of 10 mM. To isolate biotin-labelled ligation
692 junctions, 50 µl of 10 mg/ml Dynabeads MyOne Streptavidin C1 (Life Technologies, 65001)
693 were washed with 400 µl of 1× Tween washing buffer (TWB; 5 mM Tris-HCl pH 7.5, 0.5 mM
694 EDTA, 1 M NaCl, 0.05% Tween-20), collected with a magnet, resuspended in 400 µl of 2×
695 binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) and added to the sample
696 suspended in 330 µl TLE. Biotinylated DNA was bound to the beads by incubating the sample
697 for 15 min at room temperature with slow rotation. Subsequently, the DNA-bound beads were
698 captured with a magnet, washed twice with 400 µl of 1× binding buffer, washed once in 100

699 µl of 1× TLE T4 ligase buffer and resuspeded in 41 µl of TLE. For polyadenylation, 5 µl of 10×
700 NEBuffer2.1, 1 µl of 10 mM dATP and 3 µl of 5 U/µl of Klenow fragment (3′→ 5′ exo (-)) (NEB,
701 M0212) were added and the sample was incubated for 30 min at 37 °C followed by a
702 deactivation step at 65 °C for 20 min. Beads were collected with a magnet, washed once with
703 400 µl 1× Quick ligation buffer (NEB, M2200) and resuspended in 46.5 µl of 1× Quick ligation
704 buffer (NEB, M2200). 2.5 µl of DNA Quick ligase (NEB, M2200) and 0.5 µl of 50 µM annealed
705 TruSeq adapters were added and incubated for 1 h at room temperature. Beads were
706 separated on a magnet, resuspended in 400 µl of 1× TWB and washed for 5 min at room
707 temperature with rotation. Beads were washed on the magnet with 200 µl 1× binding buffer
708 and 200 µl of 1× NEBuffer2.1 and resuspended in 20 µl of 1× NEBuffer2.1. The library was
709 amplified in eight separate reactions of 50 µl. Per reaction, 1.5 µl of 25 µM TruSeq PCR primer
710 cocktail (TruSeq PCR primer cocktail_F, 5´-AATGATACGGCGACCACCGAG-3´; TruSeq
711 PCR primer cocktail_R; 5´-CAAGCAGAAGACGGCATACGAG-3´), 25 µl of 2× Kapa HiFi
712 HotStart Ready Mix (Kapa Biosystems, KR0370) and 21.5 µl of water were added to 2 µl of
713 library bound to the beads. Amplification was performed as follows: 3 min at 95 °C, 5 cycles
714 of 20 s at 98 °C, 30 s at 63 °C and 30 s at 72 °C, 1 cycle of 1 min at 72 °C, hold at 4 °C. The
715 PCR reactions were pooled and the beads were removed from the supernatant using a
716 magnet. The library was purified by addition of 1.5 volumes of Agencourt AMPure XP beads
717 (Beckman Coulter), according to the manufacturer's instructions. The sample was eluted off
718 the beads using 25 µl of 1× TLE buffer, transferred to a fresh tube and the concentration was
719 determined using Qubit (Qubit dsDNA HS Assay Kit, Thermo Fisher) and qPCR (KAPA SYBR
720 FAST qPCR Master Mix, Kapa Biosystems), according to the manufacturer's instructions.
721 Library size distributions were determined on a 5% polyacrylamide gel. Paired-end 75-bp
722 sequencing was carried out using the Illumina NextSeq 500 system with mid or high output
723 NextSeq 500/550 kits v.2.5 according to the manufacturer's instructions.
724

725 **Mapping of Hi-C reads and generation of interaction matrices.** Reads were mapped to a
726 modified version of the TbruceiLister427_2018 genome assembly (downloaded from
727 TriTrypDB, release 43) containing the following modifications. For all Hi-C experiments, we
728 masked a newly discovered misassembly in bloodstream expression site 2 (BES2) with Ns.
729 For Hi-C experiments in 2T1-control [27] and VEX2 knockdown cells, we added the transfected
730 constructs as separate contigs to the genome. The construct sequences, as well as the
731 modified genome have been deposited together with the results of the analyses under
732 https://doi.org/10.5281/zenodo.3568483. Mapping, filtering, normalization and read counting
733 were performed by the mHi-C pipeline as described in [17]. We modified the pipeline to be
734 compatible with the *T. brucei* genome assembly and also incorporated a merging step for the
735 individual replicates after the removal of duplicate reads, but before data normalization (step
736 4) in order to avoid the introduction of any bias by the merge. We chose ICEing as the
737 normalization method and finally filtered the mHi-C outcome by the posterior probability of 0.6
738 (i.e. reads are assigned to a bin with a likelihood of at least 60%). Downstream analyses such
739 as normalizing for the different ploidy within the *T. brucei* genome assembly, have been
740 implemented with in-house scripts. The digestion of the reference genome with the restriction
741 site has been implemented using HiC-Pro Utilities [47]. All scripts necessary to reproduce the
742 Hi-C analyses can be found at: https://github.com/bgbrink/PRJEB35632.
743
744 **Virtual 4C analysis.** To visualize interactions between one genomic region (viewpoint) and
745 all other genomic sites, relevant bins were extracted from a 20-kb or 50-kb Hi-C matrix. An
746 average interaction value for every genomic bin was calculated if the viewpoint regions
747 spanned more than one bin. The coordinates that define the different viewpoints used in this
748 study are shown in Data S1 sheet 2. To determine the relative interaction frequency of a
749 viewpoint with chromosome cores and sub-telomeres, the average interaction frequency of
750 the viewpoint with each chromosome core and sub-telomere was calculated based on the

relative interaction frequencies extracted by virtual 4C analysis. The ratio between the average interaction frequency (core) and the average interaction frequency (sub-telomeres) was calculated for each chromosome and plotted as one dot. The virtual 4C analysis has been implemented using HiC sunt dracones (https://doi.org/10.5281/zenodo.3570496). All scripts necessary to reproduce the Hi-C analyses can be found at: https://github.com/bgbrink/PRJEB35632.

Statistical analysis. All statistical analysis was performed using GraphPad Prism Software (version 7.0). A detailed summary of *n* and *p* values for all the analyses performed in this study is provided in Data S1 sheet 3.

Resources and Reagents. All unique materials are available on request.

Data S1. (separate file)

Sheet 1. ChIP-Seq data for VEX1^{myc} expressing bloodstream form *T. brucei*

Sheet 2. Coordinates of the viewpoints used in the virtual 4C analyses

Sheet 3. Summary of all statistical analyses

References for Methods Section:

27 Alsford, S., Kawahara, T., Glover, L. & Horn, D. Tagging a *T. brucei RRNA* locus improves stable transfection efficiency and circumvents inducible expression position effects. *Mol Biochem Parasitol* **144**, 142-148 (2005).

28 Glover, L. *et al.* Genome-scale RNAi screens for high-throughput phenotyping in bloodstream-form African trypanosomes. *Nat Protoc* **10**, 106-133 (2015).

29 Alsford, S. & Horn, D. Elongator protein 3b negatively regulates ribosomal DNA transcription in african trypanosomes. *Mol Cell Biol* **31**, 1822-1832 (2011).

30 Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416-422 (2005).

31 Becker, M. *et al.* Isolation of the repertoire of VSG expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. *Genome Res* **14**, 2319-2329 (2004).

32 Hertz-Fowler, C. *et al.* Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One* **3**, e3527 (2008).

33 Kolev, N. G., Ramey-Butler, K., Cross, G. A., Ullu, E. & Tschudi, C. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science* **338**, 1352-1353 (2012).

34 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

35 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

36 Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).

37 Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).

38 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).

39 Siegel, T. N., Hekstra, D. R. & Cross, G. A. Analysis of the *Trypanosoma brucei* cell cycle by quantitative DAPI imaging. *Mol Biochem Parasitol* **160**, 171-174 (2008).

40 Woodward, R. & Gull, K. Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of *Trypanosoma brucei*. *J Cell Sci* **95 (Pt 1)**, 49-57 (1990).

41 Park, J. H., Jensen, B. C., Kifer, C. T. & Parsons, M. A novel nucleolar G-protein conserved in eukaryotes. *J Cell Sci* **114**, 173-185 (2001).

42 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).

43 Kraus, A. J., Brink, B. G. & Siegel, T. N. Efficient and specific oligo-based depletion of rRNA. *Sci Rep-Uk* **9**, 12281 (2019).

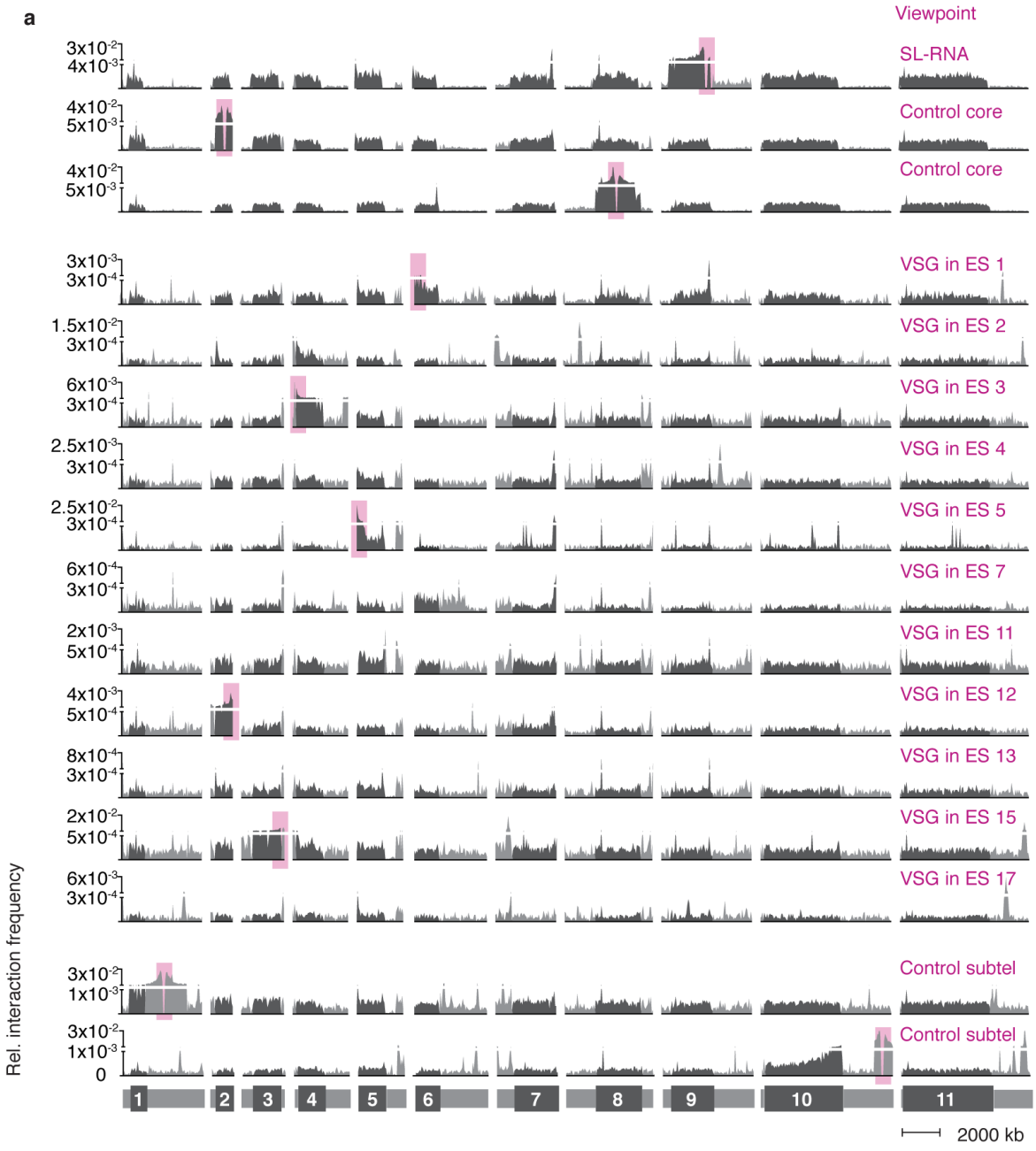
44 Aslett, M. *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* **38**, D457-D462 (2010).

45 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

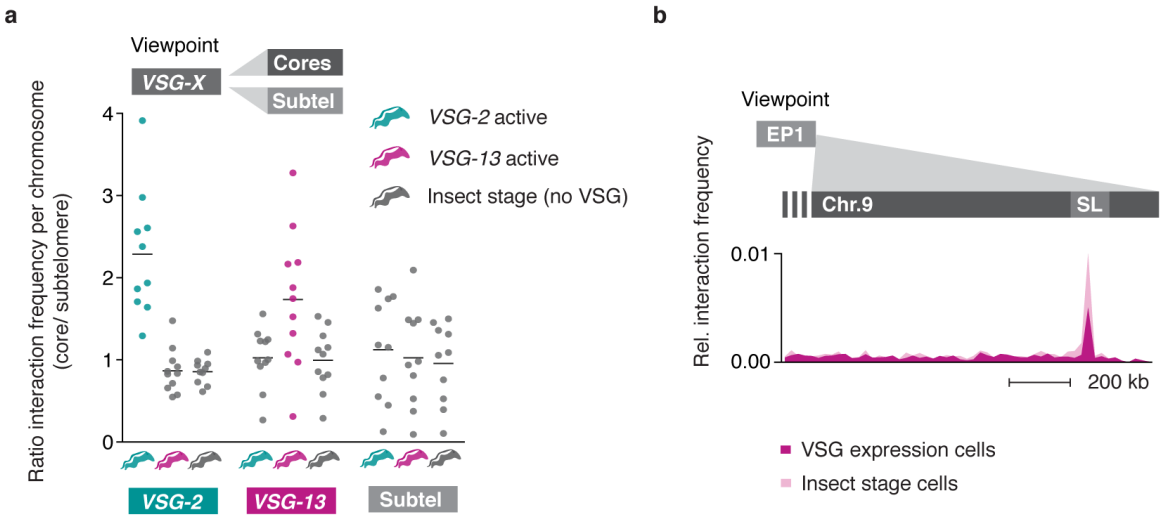
46 Wedel, C., Forstner, K. U., Derr, R. & Siegel, T. N. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *Embo J* **36**, 2581-2594 (2017).

47 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259 (2015).

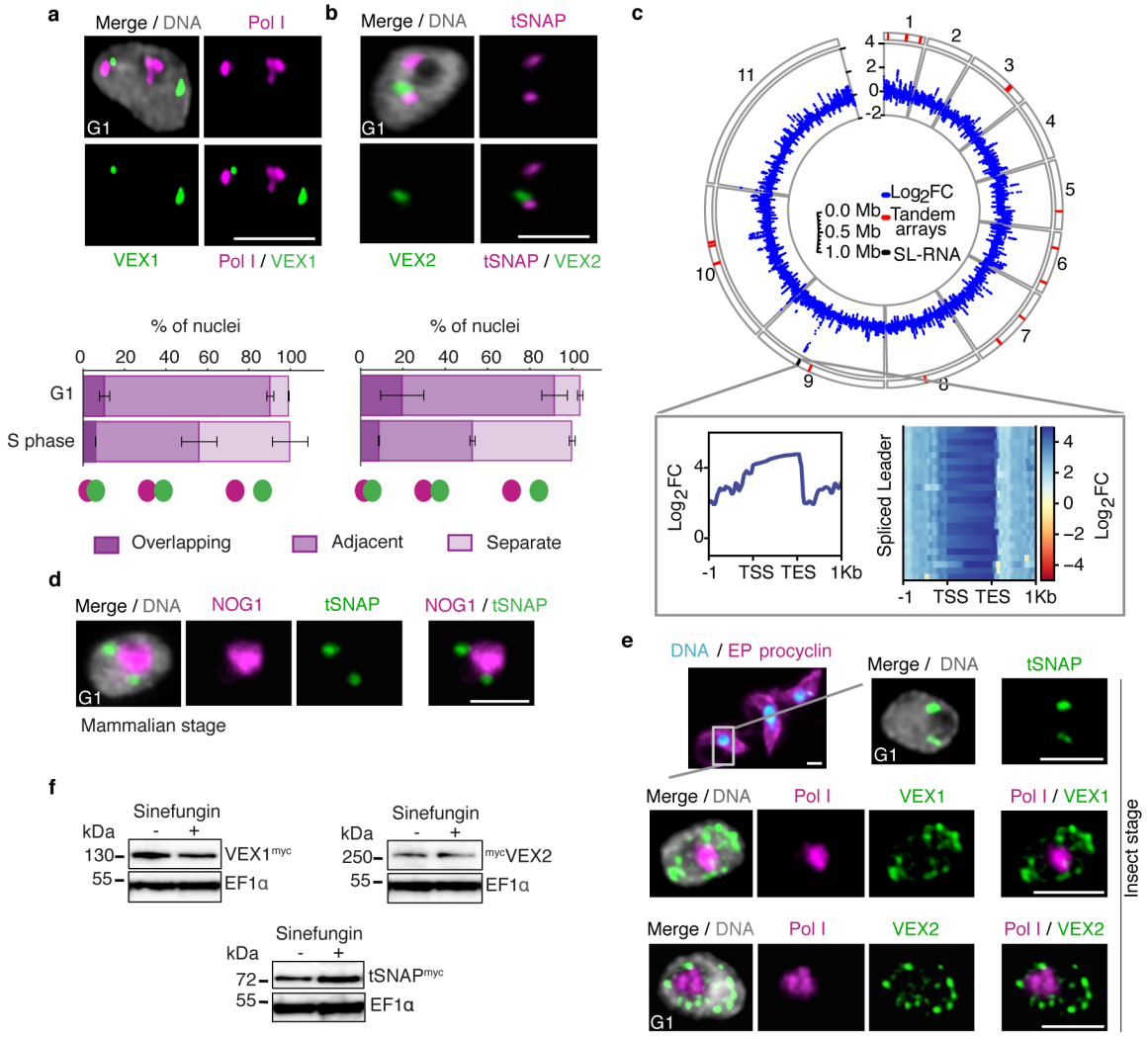
Extended Data:



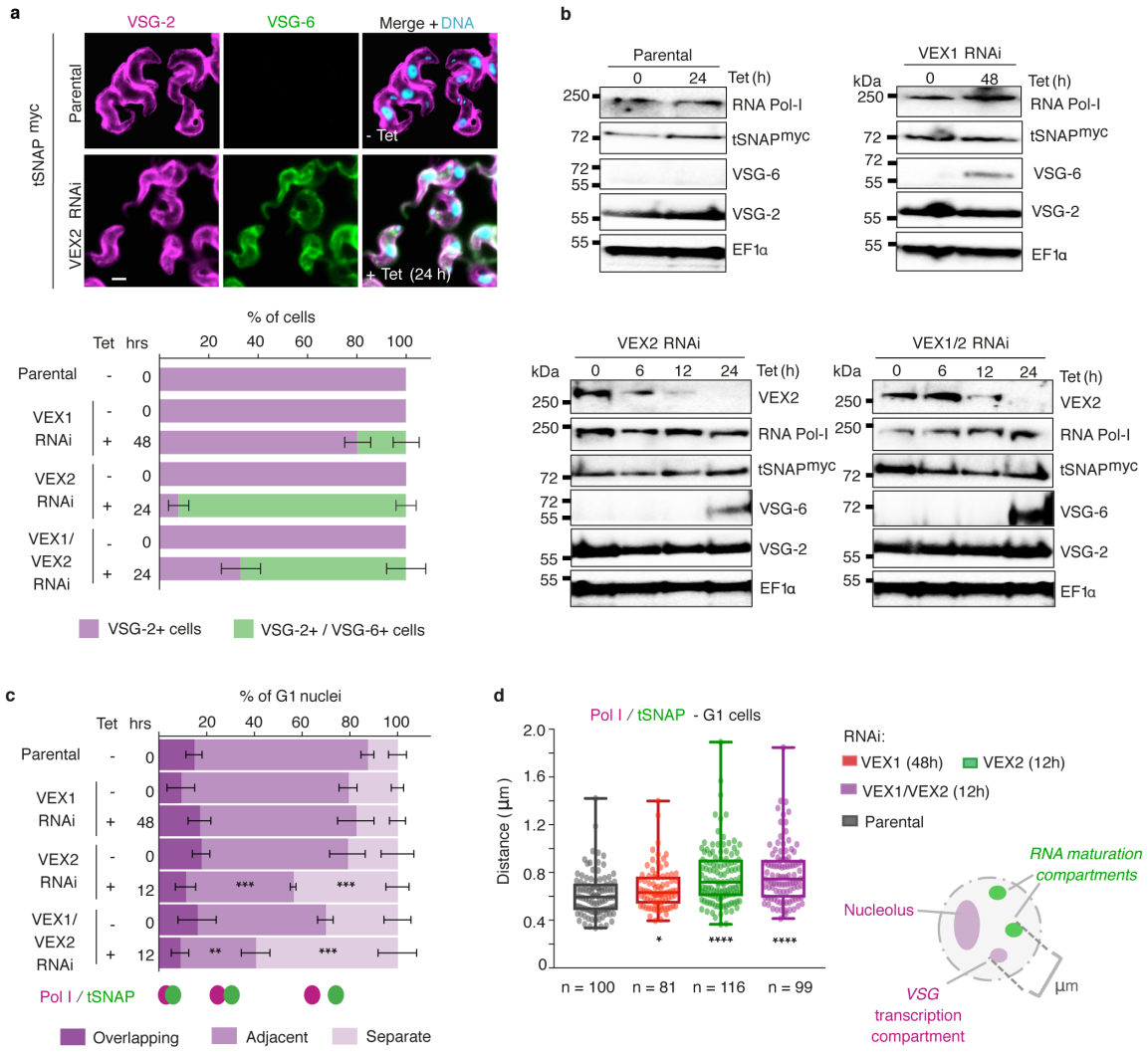
Extended Data Fig. 1 Genome-wide interaction frequencies of VSG genes in expression sites and the SL-RNA locus. a, Hi-C (virtual 4C) analysis with locations of viewpoints marked by pink boxes. Viewpoints VSG ES 4, 7, 11 and 17 are located on intermediate chromosomes that are not depicted in this figure. Interaction frequencies between each viewpoint and the 11 mega-base chromosomes are shown. Chromosome cores, dark grey; sub-telomeric regions, light grey. The hemizygous sub-telomeric regions of each chromosome are displayed in the following order: 5´(haplotype A)–5´(haplotype B)–diploid chromosome core–3´(haplotype A)–3´(haplotype B). Bin size 50 kb. Virtual 4C analyses in **a-b** are based on Hi-C experiments of *VSG-2* expressing cells (n=2, the average is shown). **b,** Virtual 4C analysis, viewpoint: SL-RNA locus (chr. 9). Relative interaction frequencies between the viewpoint and the active VSG ES 1 (cyan) and inactive VSG ESs (magenta) is shown. Each dot represents the average value for one expression site. Bin size 20 kb. **c,** Immunofluorescence-based colocalisation studies of tSNAP^{myc} (SL-RNA transcription compartment) and a nucleolar and active-*VSG* transcription compartment marker (Pol I, largest subunit) using super resolution microscopy. DNA was counter-stained with DAPI; the images correspond to maximal 3D projections of stacks of 0.1 µm slices and are representative of two biological replicates and three independent experiments; scale bars 2 µm.



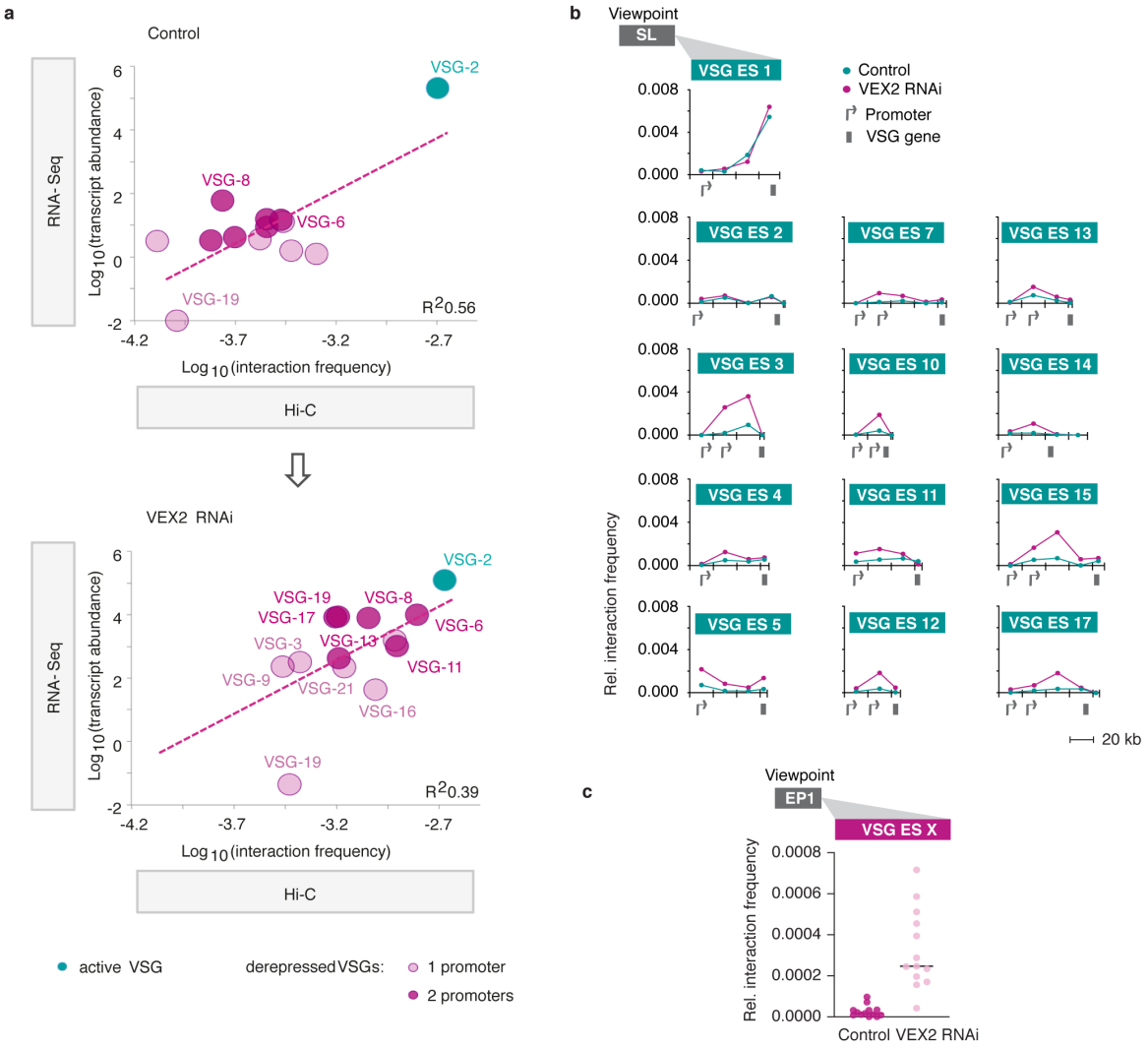
Extended Data Fig. 2 Changes in DNA-DNA interactions following a change in VSG isoform expression. a, Hi-C (virtual 4C) analysis, viewpoints *VSG-2*, *VSG-13* and a sub-telomeric control region. Each dot represents the ratio of: the average interaction frequency of the viewpoint with the chromosome core / the average interaction frequency with the sub-telomeres. One dot per chromosome is plotted. The black bar marks the median ratio per viewpoint. Bin size 50 kb. **b,** Virtual 4C analysis, viewpoint: EP1 gene array (chr. 10). Relative interaction frequencies between the EP1 array and the SL-RNA locus are plotted. Bin size 20 kb.



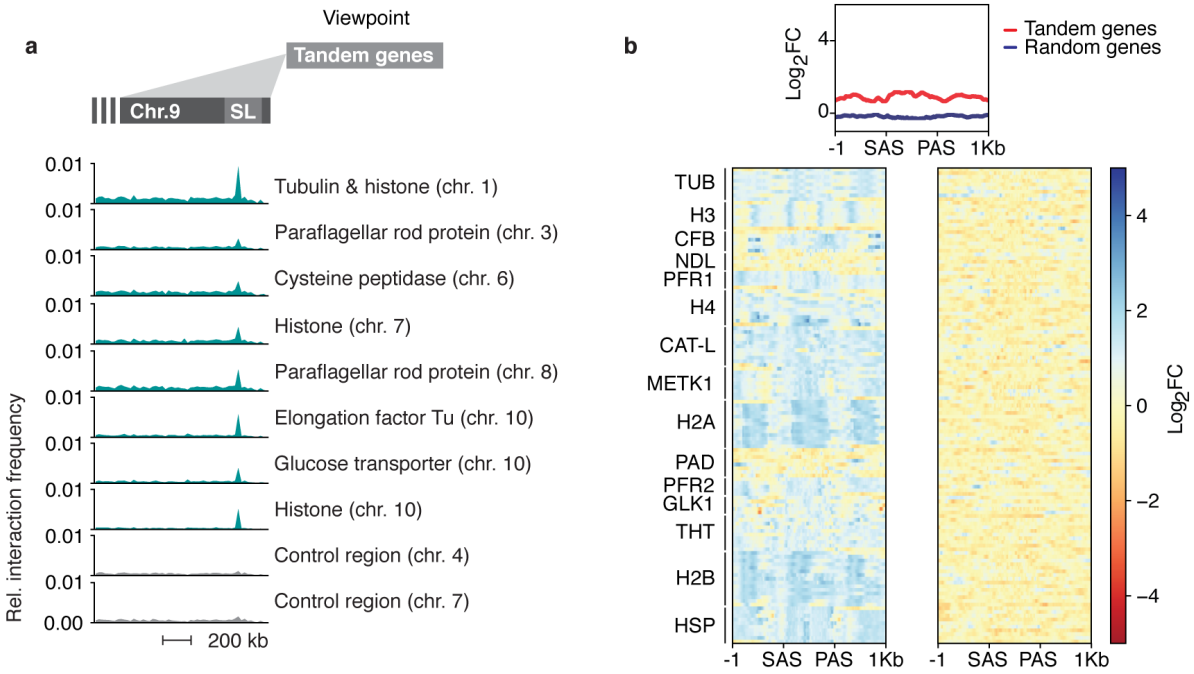
Extended Data Fig. 3. The VEX complex associates with both the active-VSG and the *Spliced Leader* (*SL*)-locus in a cell cycle and developmental stage-dependent manner. a-b, Immunofluorescence-based colocalisation studies of VEX1^{myc} / Pol I and ^{GFP}VEX2 / tSNAP^{myc} in bloodstream form cells. tSNAP and Pol I were used as markers for the SL-RNA and VSG transcription compartments, respectively. The stacked bar graphs depict proportions of nuclei with overlapping, adjacent or separate signals and values are averages of two independent experiments (≥100 nuclei for G1 and S phase cells); detailed *n* and *p* values are provided in Data S1 sheet 3. **c,** VEX1^{myc} chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) analysis. The circle plot represents log2 fold change of ChIP versus Input of non-overlapping 1 kbp bins of the 11 megabase chromosomes; outside track shows tandem arrays (red) and the SL-RNA locus (black). An inset zooming on the SL-RNA locus is depicted: metagene plot (left-hand side) and heat-map (right-hand side) of SL-gene loci. Bin size 300 bp. **d,** Immunofluorescence-based colocalisation studies of tSNAP^{myc} and a nucleolar marker (NOG1) in bloodstream forms. **e,** Localisation of tSNAP^{GFP} and colocalisation studies of VEX1^{myc} or ^{myc}VEX2 and Pol I in procyclic forms (insect-stage), using immunofluorescence. Procyclic forms do not express VSGs whereas procyclins are the major surface glycoprotein. Images in **a-b** / **d-e** were obtained with super resolution microscopy and correspond to maximal 3D projections of stacks of 0.1 µm slices; DNA was counter-stained with DAPI; scale bars 2 µm. **f,** Western-blot analysis of VEX1^{myc}, ^{myc}VEX2 and tSNAP^{myc} before and after sinefungin treatment (5 µg ml^{-1} for 30 min at 37°C), which blocks *trans*-splicing in trypanosomes. Data in **a-b** and **d-f** are representative of at least two independent biological replicates and two independent experiments.



Extended Data Fig. 4. Pol I and tSNAP expression and localisation following knockdown of the VEX complex. a, Immunofluorescence-based analysis of VSG expression following tetracycline (Tet) inducible VEX1 knockdown, VEX2 knockdown or VEX1/VEX2 knockdown. VSG-2 (magenta) is the active-*VSG* and VSG-6 (green) is a silent-*VSG* in this strain. The stacked bar graph depicts percentages of VSG-2 single positive cells and VSG-2/VSG-6 double positive cells; values are averages of two independent experiments and two biological replicates. DNA was counter-stained with DAPI; scale bar 2 µm. **b,** Western-blot analysis of VEX2, Pol I, tSNAP^{myc}, VSG-6 and VSG-2 expression following VEX1, VEX2 or VEX1/VEX2 knockdown. EF1α was used as a loading control. The data is representative of two independent experiments and two biological replicates. **c-d**, Immunofluorescence-based colocalisation studies of tSNAP^{myc} and a nucleolar and active-*VSG* marker (Pol I, large subunit). The stacked bar graph in **c** depicts proportions of G1 nuclei with tSNAP^{myc} / Pol I overlapping, adjacent or separate signals following tetracycline (Tet) inducible VEX1 (48 h), VEX2 (12 h) or VEX1/VEX2 knockdown (12 h). tSNAP^{myc} / active-*VSG* localisation were not monitored beyond 12 h following VEX2 and VEX1/2 knockdown as Pol I signal drops below detection at later time-points. The values are averages of two independent experiments and two biological replicates (≥100 G1 nuclei). In the box plot in **d**, the distance between the edges of the ESB and tSNAP foci was measured in > 81 G1 nuclei. The centre lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend from maximal to the minimal values; all data points are shown. In **a/c**, error bars, SD. In **c-d**, knockdown conditions were compared to parental cells using two-tailed paired (**c**) or unpaired (**d**) Student's *t*-tests: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. Detailed *n* and *p* values are provided in Data S1 sheet 3.



Extended Data Fig. 5. Genome-wide changes in VEX2 depleted cells. a, Correlation between the average interaction frequency of VSG expression-sites as viewpoint with the SL-RNA locus and VSG expression in reads per kilobase per million (RNA-seq data from [16]) in control and VEX2-depleted cells. **b,** Hi-C (virtual 4C) analysis, viewpoint: SL-RNA locus (chr. 9). Relative interaction frequencies between the viewpoint and VSG expression sites are shown. Bin size 20 kb. **c,** Virtual 4C analysis, viewpoint: EP1 gene array (chr.10). Relative interaction frequencies between the viewpoint and the VSG expression sites are shown. Each dot represents the average value for one expression site. Bin size 20 kb.



Extended Data Fig. 6. Tandem arrays interact at lower frequency with the *Spliced Leader* (*SL*)-locus. a, Hi-C (virtual 4C) analysis, viewpoint: different tandem gene arrays and control sites. Relative interaction frequencies between the different viewpoints and the SL-RNA locus are plotted. Bin size 20 kb. **b,** VEX1^{myc} chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) analysis. Top panel, metagene plot for tandem genes compared to randomly selected genes with no paralogues. Lower left, heat map of tandem genes. Lower right, randomly selected genes with no paralogues.